Determinants of patient survival following acute Ebola virus infection.

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy.

by

Natasha Y. Rickett (MRes)

September 2019

Abstract

The 2013-2016 outbreak of *Zaire ebolavirus* (EBOV) in West Africa was unprecedented in scale and has sparked a great deal of interest into the pathogen. EBOV is the most pathogenic species in the family *Filoviridae*. Early symptoms are non-specific and flu-like, but these rapidly dissolve into systemic issues, frequently resulting in death. Fatal infections are associated with uncontrolled inflammation, while survivors have an initial robust antiviral response followed by an effective antigen-specific response. However, there are a myriad of other factors influencing patient outcome, including viral load and co-infections. The elucidation of these factors could have major therapeutic implications.

RNA-sequencing via Illumina and Nanopore platforms was performed on the diagnostic samples of patients during the outbreak in order to explore many facets of EBOV biology. The nature of these samples allowed investigation into the determinants of patient outcome following Ebola virus disease (EVD). Blood was taken from acutely ill individuals at a time when it was not known whether they would survive and clear the infection or succumb to it. These data allowed investigation into viral genome variation and the potential role of viral quasispecies in patient outcome, as well as the study of defective interfering particles *in vitro* – a potential mechanism for viral persistence. The immune response to the pathogen was assessed via the comparison of gene expression and predicted immune cell profiles. Bioinformatic techniques were also employed to search these diagnostic samples for other microorganism such as bacterial and fungal species.

Here we demonstrate that higher viral load and malaria co-infection are correlated with poor patient outcome. There are a number of genes that are differentially expressed according to patient outcome – e.g. those associated with blood clotting and complement regulation. These data allowed the employment of a linear discriminant analysis which, following qRT-PCR of EBOV and three host genes, was able to predict patient outcome with exceptional accuracy. Thus suggesting that the

response of survivors differs to that of fatalities during EBOV infection. The identification of potential microorganisms in the blood of EVD patients is suggestive of gut translocation of bacteria. This points towards the potential use of antibacterial agents in conjunction with treatment for EVD during outbreak settings. Potential defective interfering particles were readily detected upon sequencing of infected THP-1 cells, however, were not often found in clinical samples taken from EVD patients. These samples included those from immune-privileged niches, i.e. semen, and may suggest that Dis are not responsible for the persistence of EBOV following acute infection.

This thesis highlights the fact that animal studies, cell culture and the analysis of clinical samples should be considered in tandem in order to effectively study all aspects of EBOV biology.

Acknowledgements

This research was funded by the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Emerging and Zoonotic Infections at the University of Liverpool in partnership with Public Health England.

I'd like to extend my eternal thanks to my supervisor Julian Hiscox for being exceptionally supportive and always pushing me to get the most out of this PhD. Also to the entire Hiscox laboratory for making Liverpool a wonderful place to be. Thanks also to Miles Carroll as well as Roger Hewson, David Matthews and Neil Hall. This work could not have been completed without the assistance of the skilled individuals at Porton Down, Public Health England and the Centre of Genomic Research in Liverpool. Furthermore, thank you to Sophie Duraffour and Stephan Günter for facilitating my training in handling Ebola and travel to West Africa during the outbreak with EMLab.

A special thank you to Catherine McCarthy, Victoria Wood and my other friends who have kept me going with pints, sushi and escape rooms. Thanks to my Dad who provided great advice as well as innumerable cups of tea.

This is dedicated to my Mum – forever my inspiration.

Contents

List of figures

List of tables

List of abbreviations

ADAR	Adenosine Deaminases Acting on RNA
BDBV	Bundibugyo ebolavirus
BOMV	Bombali virus
bp	Base pairs
C_D	Community deaths
CDC	Centers for Disease Control and Prevention
CFR	Case fatality rate
CL4	Containment level 4
CMV	Cytomegalovirus
CPM	Counts per million
Ct	Cycle threshold
DC	Dendritic cell
DCQ	Digital cell quantification
DE	Differentially expressed
DI	Defective interfering
DNA	Deoxyribonucleic acid
EBOV	Zaire ebolavirus
EBV	Epstein Barr virus
EMLab	European Mobile Laboratory
ETC	Ebola treatment centre
EVD	Ebola virus disease
FACS	Fluorescence-activated cell sorting
FDR	False discovery rate
FPKM	Fragments Per Kilobase Million
GBV-C	GB virus C, formerly hepatitis G virus
GLM	General linear model
GP	Glycoprotein
h.p.i	Hours post infection
H_F	Hospitalised fatalities
H_S	Hospitalised survivors
HBV	Hepatitis B virus
HCV	Hepatitis C virus
HHV4	Human herpes virus 4, formerly Epstein-Barr virus
HIV	Human immunodeficiency virus
HPV	Human papillomavirus
HSE	Herpes simplex encephalitis
HSV	Herpes simplex virus
IFIT	IFN-induced protein with tetratricopeptide repeats
IFN	Interferon

IL	Interleukin
IPA	Ingenuity Pathway Analysis
IRF	Interferon regulatory factor
ISG	IFN stimulated gene
L	Viral polymerase
LDA	Linear discriminant analysis
MARV	Marburg virus
MDS	Multidimensional scaling plot
MIP	Macrophage inflammatory protein
MOI	Multiplicity of infection
mRNA	Messenger RNA
MV	Minor variants
NGS	Next generation sequencing
NHP	Non-human primates
NK	Natural killer
NP	Nucleoprotein
ORF	Open reading frame
qRT-PCR	Quantitative real-time polymerase chain reaction
RDT	Rapid diagnostic test
RESTV	Reston ebolavirus
RNA	Ribonucleic acid
RNP	Ribonucleoprotein complex
RSV	Respiratory syncytial virus
SD	Standard deviation
SNP	Single nucleotide polymorphism
SUDV	Sudan ebolavirus
TAFV	Taï Forest ebolavirus
TLR	Toll-like receptor
TNF	Tumor necrosis factor
VP	Viral protein
WHO	World Health Organisation

Chapter 1: Introduction

1.1 Project overview

This project was formed in response to the beginning of the largest Ebola outbreak in history, first confirmed in West Africa in 2014. Although vast and significant improvements have been made both in terms of vaccines to *Zaire ebolavirus* (EBOV) and effective treatments in response to that outbreak, the zoonotic nature of this virus means that outbreaks are likely to occur in the future. These outbreaks will likely occur in settings where the delivery of vaccines and drugs pose challenges. Therefore, the further understanding of disease progression and pathogenesis could provide valuable information to healthcare workers, either for personalised treatments or the triage of patients. In order to explore this, three aspects were investigated: (1) The virus itself, (2) the host immune response to Ebola virus disease (EVD) and (3) the potential role of co-infections in patient outcome. Understanding the differences between acute sufferers of EVD and their fatal counterparts in terms of immune response could also potentially lead to the development of a prognostic tool. For this to be effective it would have to be able to be performed in relatively low-resource laboratories, which limited the potential methodologies that could be implemented.

Currently, a measurement of EBOV Ct (following quantitative (q)RT-PCR of blood) is used as a proxy for viral load and this gives some indication as to the prognosis of the patient. However, there are many cases when this value is intermediate and provides

no information concerning the patient's likely outcome. Once developed, such a prognostic tool may be implemented in other outbreaks, where viral load is not such an accurate indicator of outcome.

1.2 Ebola virus

Ebolavirus is a genus of the family Filoviridae and order Mononegavirales. Within this genus exists six known species: Bundibugyo ebolavirus (BDBV), Reston ebolavirus (RESTV), Taï Forest ebolavirus (TAFV), Sudan ebolavirus (SUDV), Bombali virus (BOMV) and Zaire ebolavirus (EBOV), which will be the focus of this thesis. Many of these species are capable of causing EVD, with varying mortality rates. EBOV causes the most frequent outbreaks and those capable of causing the highest case fatality rates - up to 90% (Formenty *et al.*, 2003; Weyer *et al.*, 2015). EBOV is a zoonotic virus and able to infect not only in non-human primates (NHPs) but also other mammals with bats being a likely reservoir (Leroy *et al.*, 2005; Pourrut *et al.*, 2009; Hayman *et al.*, 2010; Goldstein *et al.*, 2018). This zoonotic nature allows periodic outbreaks to occur, with the bushmeat of bats being the source of the 2013-2016 outbreak (Mari Saez *et al.*, 2015). Since their discovery, the zoonotic potential of filoviruses has been understood. Marburg virus (MARV) was confirmed to be present in *Rousettus aegypticus* bats, implicating them early on as being a potential reservoir. Since then, Ebola viruses have been confirmed to be present in bats and rodents as well as the bodies of chimpanzees surrounding outbreaks of EVD (Breman *et al.*, 1980; Leroy *et al.*, 2005).

1.2.1 Transmission

Ebolaviruses are highly transmissible, principally by direct contact with contaminated body fluids and fomites (Bausch *et al.*, 2007), though aerosol transmission has been achieved in laboratory conditions (Johnson *et al.*, 1995; Zumbrun *et al.*, 2012; Nfon *et al.*, 2013). The 2013-2016 outbreak was sparked by one zoonotic jump, with all subsequent transmissions being due to human-human interactions. This was partially facilitated by the fact that individuals continue to secrete the virus after death, and a number of traditional burial practises include direct contact with the body. One aspect of the response to outbreaks of EVD has been the dissemination of information concerning safe funeral practises (Tiffany *et al.*, 2017). The 2013-2016 outbreak also saw the first confirmed case of sexual transmission of the virus (Diallo *et al.*, 2016).

Outbreaks in human populations tend to start following an initial spillover event from an infected reservoir to an alternate species, with numerous epizootic outbreaks having occurred in NHPs in the Republic of Congo and Gabon (Leroy *et al.*, 2004). These spillover events can be due to human activity – e.g. hunting or poaching or NHPs, with infected NHPs believed to be responsible for sparking several EVD outbreaks (Formenty, Boesch, *et al.*, 1999; Formenty, Hatz, *et al.*, 1999; Reed *et al.*, 2014). This interplay between suspected reservoirs, vulnerable NHPs and human populations is not well understood when it comes to EVD.

1.2.2 Outbreaks

Since its first discovery in 1976, there have been a number of outbreaks involving ebolavirus infections, as summarised in Table 1.1. In order to confirm such outbreaks, access to adequate testing facilities is essential, as confirming EVD clinically poses

many challenges due to the non-specific nature of the exhibited symptoms. This was shown once again in the 2013-2016 outbreak in West Africa, where the initial cases were assumed to be malaria infections (Baize *et al.*, 2014).

Table 1.1: Summary of confirmed EVD outbreaks throughout history. *The location and date of outbreak is shown along with the absolute number of cases, deaths and case fatality rate (CFR). * denotes estimated CFR due to ongoing outbreak. Data sourced from Centers for Disease Control and Prevention, accessed 01/08/19* (CDC, 2019a).

Date	**Country**	**Human cases**	**Human deaths (CFR)**
Ebola virus (EBOV)			
Aug 2018 – present	Democratic Republic of the Congo Limited cases: Uganda	3 373	1 986 (59%*)
May – July 2018	Democratic Republic of the Congo	54	33 (61%)
Aug – Nov 2014	Democratic Republic of the Congo	66	49 (74%)
Dec 2013 – Jan 2016	Liberia, Sierra Leone, Guinea Limited cases: Nigeria, Mali, USA, Senegal, Spain, UK, Italy	28 646	11 323 (70-71%)
Dec 2008 – Feb 2009	Democratic Republic of the Congo	32	14 (45%)
Aug – Nov 2007	Democratic Republic of the Congo	264	187 (71%)
2005	Republic of the Congo	12	10 (83%)
Nov – Dec 2003	Republic of the Congo	35	29 (83%)
Dec 2002 – Apr 2003	Republic of the Congo	143	128 (90%)
Oct 2001 – Jul 2002	Gabon, Republic of Congo	124	97 (78%)
July 1996 – Mar 1997	Gabon	60	45 (75%)
Jan – Apr 1996	Gabon	31	21 (68%)
May – July 1995	Zaire	315	254 (81%)
Dec 1994 – Feb 1995	Gabon	52	32 (60%)
Aug 1976	Zaire	318	280 (88%)

1.2.3 2013-2016 Outbreak of EVD in West Africa

Periodic outbreaks of EVD occurred from its discovery in 1976 until the re-emergence of the virus in West Africa in 2013. This outbreak was unprecedented in scale and has elucidated previously unappreciated aspects of EBOV biology, such as persistence in semen and ocular fluid (Deen *et al.*, 2015; Varkey *et al.*, 2015). The first cases of EVD were reported in March 2014 in the forested region of Guinea (Baize *et al.*, 2014). Using phylogenetic analysis, the first cases of the outbreak were traced back to December 2013, which was supported by epidemiological evidence. These data suggested that EBOV-Makona, titled as such after a river running through Sierra Leone and subsequently along the borders of Guinea and Liberia (Kuhn *et al.*, 2014), was exposed to humans via the interaction between a young boy and a dead bat (Simon-Lorière *et al.*, 2015). Throughout this outbreak over 28,000 people were infected and 11,300 succumbed to the effects of EVD (Table 1.1). This outbreak mainly affected three countries: Guinea, Sierra Leone and Liberia, with sporadic but isolated cases occurring in surrounding countries. EBOV-Makona has been shown to have different growth kinetics when compared to other strains according to experimentally infected NHPs and mice, where delayed onset of disease and reduced lethality were observed in each model respectively (Marzi *et al.*, 2015; Smither *et al.*, 2016).

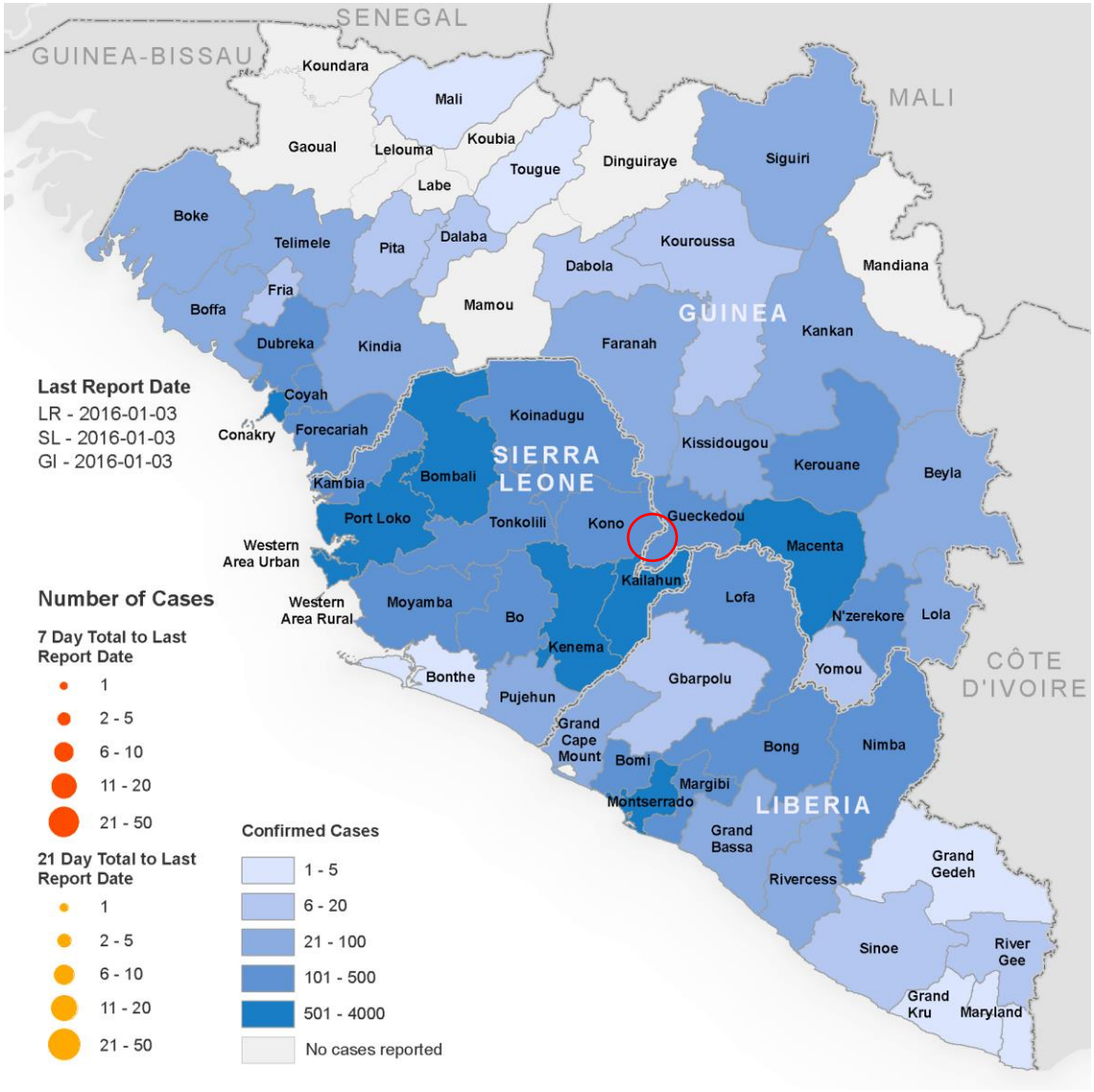


Figure 1.1: Map of West Africa during the 2013-2016 Ebola outbreak. *Shown here are Guinea, Liberia and Sierra Leone, the three most prominently affected countries of the outbreak. Districts are coloured according to number of EVD cases reported between March 2014 and March 2016. The area considered to be the epicentre of the outbreak is depicted by the red circle. The map of Ebola cases in West Africa was adapted from the World Health Organisation, accessed 04/03/19* (WHO, 2016).

The epicentre of this outbreak was critical in understanding its subsequent spread and scale. It began in a remote part of Guinea, close to the borders of Sierra Leone and Liberia (Figure 1.1). These borders are porous, and much is shared across them including entire families. This made isolation of hot zones and controlling the outbreak challenging. Interestingly, the CFR was not constant between these three majorly affected countries (Figure 1.2). This thesis focussed on samples from Guinea, the country that experienced the highest case fatality rate, and therefore potentially the most extreme conditions both for the virus and host. Guinea is a developing country with limited infrastructure. Therefore, few palliative care measures were implemented during the outbreak and even fewer opportunities to provide antivirals and/or vaccines.

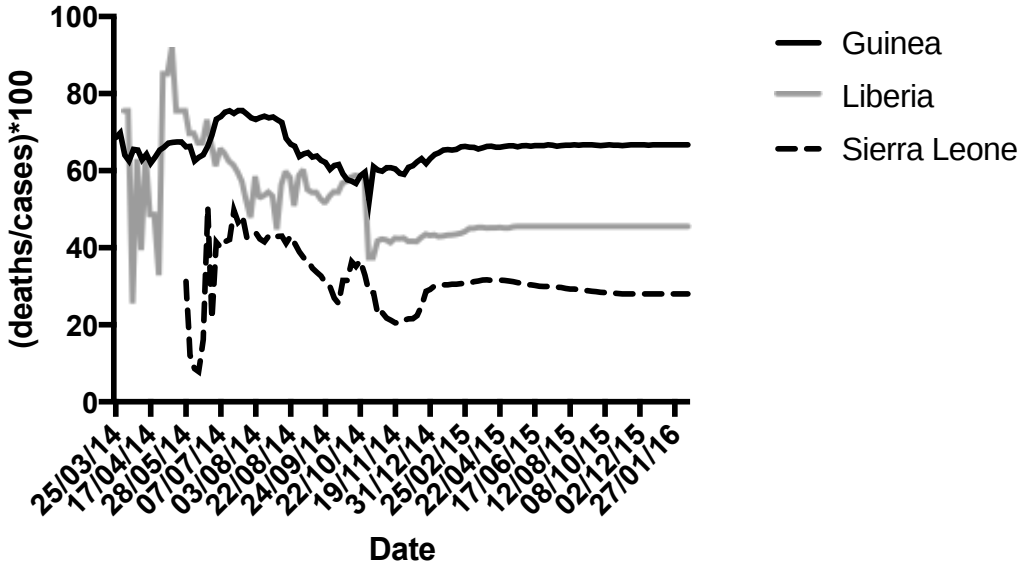


Figure 1.2: Tracking the case fatality rate across the outbreak in Guinea, Liberia and Sierra Leone. *Data collected from the Centers for Disease Control and Prevention, accessed 07/05/19* (CDC, 2019b). *Graph shows the percentage of cumulative suspected cases that have resulted in death. These are suspected and lab-confirmed cases commencing after confirmation of the outbreak in March, 2014.*

1.3 Virology

1.3.1 Virion structure

Ebolavirus virions are filamentous in their morphology, varying in length from 800 nm to 1 μm, but with a relatively uniform diameter of 60 nm (Baskerville *et al*., 1985; Noda *et al.*, 2002). The main structural proteins of EBOV are the membrane protein (VP40) and the glycoprotein (GP) (Figure 1.3A). There are also several non-structural proteins that aid in viral replication and viral transcription: i.e. VP35 and VP30, while VP24 plays a key role in viral packaging and interacting with host restriction factors. The viral polymerase (L) is critical in viral replication and transcription and the nucleoprotein (NP) encapsulates viral RNA to form the ribonucleoprotein complex (RNP) (Muhlberger *et al.*, 1999). This packaging facilitates viral replication and also protects the viral RNA when a host first becomes infected.

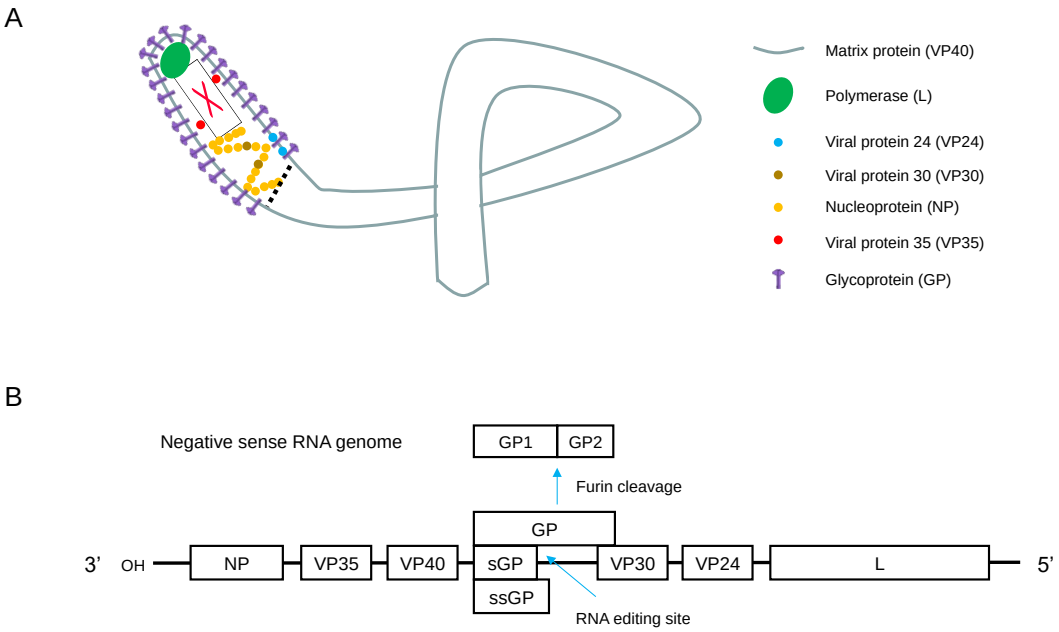


Figure 1.3: Schematic depictions of Ebolavirus and its genome. *Morphological representation of Ebola virus shows its filamentous nature, hence the genus name Filovirus (A). The outer membrane consists of interlinking matrix proteins (VP40), in which the glycoproteins (GP) are embedded, covering the entire surface area of the virion. Viral protein (VP)24 is associated with the membrane, being partially within the nucleoprotein (NP). NP is found within the capsid, along with the viral RNA to form the ribonucleoprotein (RNP) complex. Also present are the polymerase (L), replication activator (VP30) and transcriptional activator (VP35). Negative sense RNA genome of Ebola virus organised 3' to 5' (B). Like all ebolaviruses this is an unsegmented negative sense RNA genome. Transcription occurs sequentially from 3' to 5' ends. During replication, the entire genome is transcribed to produce an anti-genome, which acts as a template for the generation of further viral genomes. GP consists of GP1 and GP1, cleaved by the host protein furin. The RNA editing site allowing for the different GP products is also highlighted.*

1.3.2 Genome structure and organisation

The genome of EBOV is 18.9Kb - negative sense RNA genome encoding seven genes (Figure 1.3B) (Brauburger *et al*., 2014). Transcription is sequential from the 3' end, with all but one gene encoding for a single protein: NP, VP30, VP35, GP, VP40, VP24 and L. As well as full-length GP, this gene produces three transcriptional products: secretory (s)GP, small (s)sGP and delta-peptide (Mehedi *et al*., 2011; He *et al*., 2017; Wahl-Jensen *et al*., 2005). Flanking each EBOV gene exists an open reading frame (ORF) consisting of untranslated sequences as well as highly conserved transcription start and stop signals.

1.2.4 Viral proteins

This genome encodes seven structural proteins: NP, VP35, VP40, GP, VP30, VP24 and L as well as two non-structural proteins: sGP and ssGP (Elliott 1985, Feldman 1993).

1.2.4.1 Nucleoproteins

The major nucleoprotein, NP, and minor nucleoprotein, VP30, bind to viral RNA. NP binds to viral RNA in order to form the RNP, encapsulating the RNA, shielding it from degradation by cellular processes and facilitating viral transcription (Noda *et al*., 2010). NP is essential for viral replication and, as such, many regions of this gene are highly conserved among the order mononegavirales (Muhlburger *et al*., 1999; Muhlburger 2007; Sanchez *et al.,* 1992).

The role of VP30 is in the activation of transcription, which again is imperative for the virus lifecycle (Martinez *et al*., 2008; Muhlberger *et al*., 1999; Biedenkopf *et al*., 2016).

The activity of VP30 is enhanced by a secondary RNA structure which facilitates the production of viral transcripts via polymerase activity (Biedenkopf *et al*., 2016).

1.2.4.2 Viral protein 35 (VP35)

VP35 is a phosphoprotein with two main functions: acting as a co-factor for the viral polymerase L and as a facilitator of immune evasion. This is achieved by interfering with the host's establishment of an antiviral state. VP35 interrupts signalling pathways such as RIG-I and interferon regulatory factor (IRF-)3 (Hartman et al., 2008; Basler *et al*., 2003; Ilinykh *et al*., 2015).

1.2.4.3 Glycoprotein (GP)

The major product of the GP gene is a soluble non-structural glycoprotein – sGP – but structural proteins GP1 and GP2 can be produced following RNA editing. GP2 enables entry into host cells by acting as a fusion protein (Noda *et al*., 2002; Carette *et al*., 2011). GP1 and GP2 form a trimer in order to mediate membrane fusion (Lee *et al*., 2008). The cellular receptors for EBOV are not fully understood, but DC-SIGN, LSECtin, β-integrins and Tyro3 receptors have been implicated (Takada *et al*., 2004; Marzi *et al*., 2004; Zhao *et al*., 2016., Schornberg *et al*., 2009; Hunt *et al*., 2011). However, these factors are not essential for EBOV entry. This entry occurs via endocytosis and to permit release into the cytoplasm, GP must be catalysed by Cathepsin B/L, which trims GP1 and removes the mucin domain, allowing access to other binding sites for additional receptors in the endosome (Aleksandrowicz *et al*., 2011). GP then binds NCP1, enabled by furin (Carette *et al*., 2011). NCP1 is involved in cholesterol trafficking and found on the inner membrane of the endosome (Garver *et al*., 2002).

1.3.3.4 sGP

The primary product of the GP gene is a non-structural secretory glycoprotein – sGP. The N-terminal regions of sGP, ssGP and GP1 are identical though the C-terminals are different. Therefore, it has been hypothesised that sGP may act as a protective decoy, delaying the success of antibodies produced in response to the virus (Lee *et al*., 2008; Iwasa *et al*., 2011). sGP has also been implicated in the activation of endothelial cells and therefore promoting vascular leakage via the weakening of cell junctions (Wahl-Jensen *et al*., 2005). This is readily observed in patients with EVD. A truncated version of sGP, ssGP, can be formed, though its function is not well understood (Mehedi *et al*., 2011).

1.2.4.4 Matrix protein VP40

VP40 oligomerises in order to form stable structures, particularly in areas known as lipid rafts, at the plasma membrane (Hoenen *et al*., 2010). These areas comprise of high cholesterol and sphingolipid concentration, which, following budding, are involved in the formation of the viral envelope. VP40 forms a complex with the RNP in these areas to initiate virion self-assembly, which precedes viral budding through the membrane (Noda *et al*., 2002; Hoenen *et al*., 2010). VP40 also binds to the C-terminal of GP to act as an anchor while the mucin domain penetrates the surrounding envelope.

1.2.4.5 Viral protein 24 (VP24)

The roles of VP24 are numerous and multi-faceted. VP24 acts as a co-factor in virion assembly by associating with the RNP and potentially aiding in the packaging of NP (Watt *et al*., 2014). In the absence of VP24, virion packaging efficacy is interrupted,

though the viral lifecycle is unaffected (Zhang *et al*., 2012; Watt *et al*., 2014). The other main role of VP24 is that of an interferon antagonist (Zhang *et al*., 2012). The binding of VP24 to karyopherin-α halts the translocation of STAT1/STAT2, which spark the transcription of IFN response genes. In experimentally infected guinea pigs, mutations in the VP24 region were associated with an increase in pathogenicity as the virus adapted to the guinea pig host (Dowall *et al.*, 2014).

1.2.4.6 Viral polymerase (L)

The RNA dependent RNA polymerase (RdRp) associates with viral RNA in order to transcribe messenger (m)RNA and promote replication. The polymerase binds to the 3' domain in a specific region in the UTR (Muhlberger *et al*., 2007). The concentration of NP will dictate whether the polymerase will produce full-length RNA anti-genomes or mRNA for viral genes by association with VP30 (Noda *et al*., 2011). This gene is highly conserved across the *Ebolavirus* genus due to its essential role in the replication and spread of the virus (Oany *et al.*, 2015).

1.2.5 Viral replication

EBOV replication occurs within inclusion bodies formed between NP and host proteins (Hoenen *et al*., 2012). Once formed, this virion can enter endosomes by micropinocytosis or clathrin-mediated endocytosis. Both processes involve the formation of invaginations in the plasma membrane that engulf the endosome, in which GP plays a role. NCP1 is able to initiate the fusion of the viral membrane to the vesicle membrane, thus releasing the RNP into the cytoplasm. The critical role of GP in viral entry has been demonstrated via the introduction of mutations into GP *in vitro*, where these changes had the ability to restrict viral entry (Martinez *et al.*, 2013).

Following the release of the RNP, sequential transcription of viral mRNAs occurs, along with genome replication. This process begins with the synthesis of a complementary antigenomic RNA, regulated by L, which acts as a template for future replication. The trigger for replication is thought to be initiated by the expression of NP, once encapsidation of the antigenomic template has occurred. Upon the generation of sufficient viral material within the inclusion bodies, these RNA genomes are encapsidated and virions formed. Here, RNP complexes they migrate to VP40 assemblies at the plasma membrane (Nanbo *et al.*, 2013).

1.2.6 The formation of viral quasispecies

The error-prone nature of the viral polymerase, along with potential recombination events, leads to the generation of viral quasispecies. These describe the mutant spectra of closely related viral genomes that exist in one infection, constantly undergoing genetic variation and competition with other variants (Domingo *et al.*, 2012). These viral populations exist even in a single cell as well as at the level of an infected organism (Del Portillo *et al.*, 2011; McWilliam Leitch and McLauchlan, 2013; Gire *et al.*, 2014). This allows viral genomes to shift and gain or lose fitness throughout the course of an infection, which has been demonstrated to be critical in viral adaptation (Domingo and Holland, 1997). The association between viral fitness, disease progression and response to antiviral therapy was demonstrated in hepatitis C virus (HCV) (Farci *et al.*, 2000, 2006).

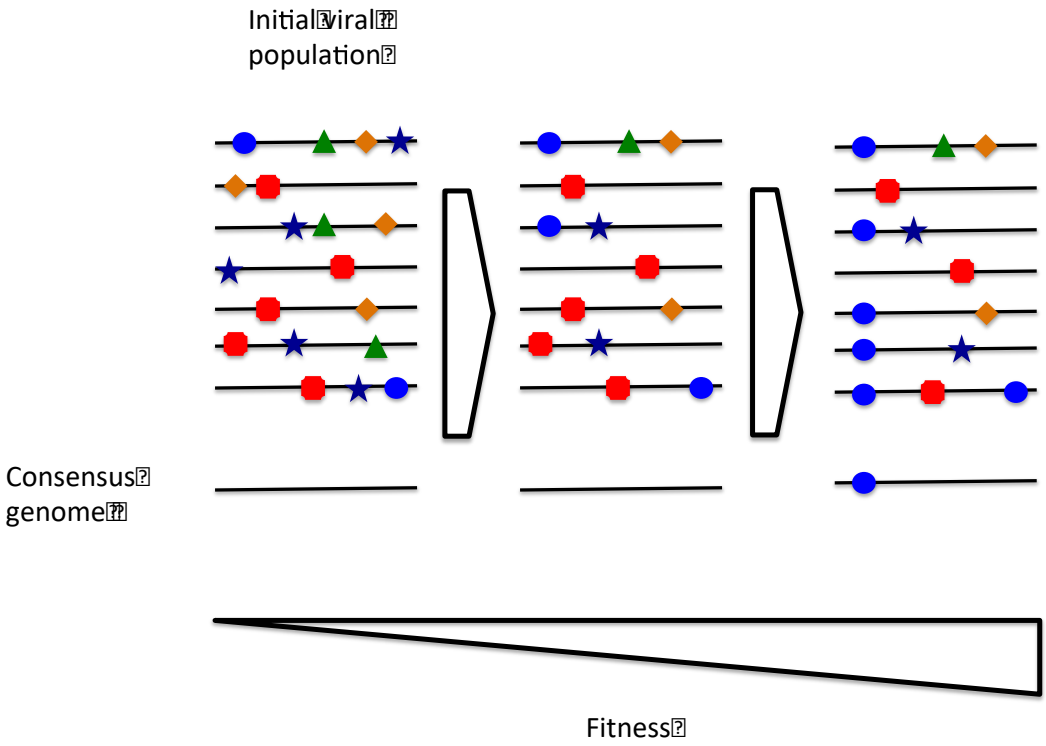


Figure 1.4: Schematic diagram depicting the selection for certain mutations in the evolution of viral quasispecies. *The error-prone nature of the viral polymerase allows the introduction for point mutations. The initial viral population is subject to change over time. Those mutations that provide a fitness benefit are maintained within the viral quasispecies, whilst those that are detrimental are selected against. Therefore, over time an increase in fitness is expected, along with these accumulated mutations. Here, the viral genome is depicted as a horizontal black line, while mutations are represented by coloured shapes.*

1.4 Persistent filovirus infections

Due to the sheer number of infected individuals, the outbreak in West Africa highlighted a number of aspects of EBOV that were previously unknown or poorly understood. One example of this is the persistence of the virus for long periods of time within an infected host, often not accompanied by physiological symptoms (Leroy *et al.*, 2001). This persistence, especially in body compartments considered to be immune-privileged niches such as the testes, are particularly problematic as they raise the potential for further human-human transmissions via sexual contact, as shown in EBOV in 2015 (Christie *et al.*, 2015). The first case of sexual transmissions of a filovirus was observed in an outbreak of Marburg virus (MARV) in 1968 (Martini and Schmidt, 1968). Additionally, such persistence could be the cause of substantial sequalae observed in survivors, among whom there are a plethora of long-term health consequences associated with EVD (Clark *et al.*, 2012; Quaglio *et al.*, 2019; Ficenec *et al.*, 2019; Kelly *et al.*, 2019). This includes damage to reproductive health, psychological effects and injury to the sight of convalescent patients.

1.4.1 Defective interfering (DI) particles

In most RNA viruses, truncated forms of the viral genome are produced during replication during growth *in vitro* (Calain *et al.*, 1999; Rima and Duprex, 2005). These defective genomes have been suggested as antiviral in nature, due to their ability to interfere with viral transcription and/or packaging (Marriott and Dimmock, 2010). However, the understanding of the role of DIs *in vivo* is less well understood (Barrett *et al.*, 1986). Broadly speaking, DIs fall into three categories: deletion, copy-back and hairpin DIs (Dimmock and Easton, 2014). This self-imposed system has been suggested

as the mechanism by which EBOV, and other RNA viruses, maintain persistent infections. Such infections appear to lack visible symptoms and represent infections in which the environment has become immune-tolerant to the pathogen. Viral quiescence describes a latent form of the virus, which is able to sequester and remain in the infected host for extended periods of time, as demonstrated in many viruses including HSV, CMV and Rift Valley fever virus (Jaimieson *et al.,* 1995; van Vuren *et al*., 2019; Russell and Tscharke, 2016; Poole and Sinclair, 2015).

1.5 Pathology and the immune response

1.5.1 Pathogenesis

The high CFR observed in outbreaks of EVD is achieved by "immune paralysis" – a total overwhelming of the immune system due to uncontrolled viral replication. Analysis of patients from the 2013-2016 EBOV outbreak in West Africa including four patients who received intensive treatment at Emory University Hospital (GA, USA) indicated that infected individuals had a robust immune response during the acute phase of EVD (McElroy *et al*., 2015). This, and findings from previous outbreaks, challenges previous *in vitro* studies that suggested EBOV was able to effectively suppress the immune system (McElroy *et al*., 2015). In these studies, EBOV appeared to diminish the activity of T lymphocytes, dendritic cells (DCs) and macrophages. These cell types are targeted by EBOV early in infection (Geisbert, *et al.*, 2003).

1.5.1 Inflammatory response

Certainly, in NHP models of fatal EBOV infection, an extreme aberrant immunological status and anti-inflammatory response was shown to contribute to the development of fatal haemorrhagic fever (Ebihara *et al*., 2011) and this has also been observed together with lymphocyte apoptosis and other indicators of immunopathology in fatal human infections with EVD (Wauquier *et al.*, 2010). Analysis of samples from EBOV-infected patients treated in Guinea also revealed an immune component influenced survival (McElroy *et al*., 2015; Ruibal *et al*., 2016). The study of patient samples taken from previous outbreaks suggests that host responses may delineate survival and fatal outcomes and potential biomarkers indicative of these outcomes can be identified. In the 2000–2001 Sudan-associated Ebola (SUDV) outbreak in Uganda, the case fatality rate for paediatric patients was lower than for adults (Mupere *et al*., 2001). Data indicated that paediatric patients who survived had differential abundance of certain serum proteins from paediatric patients who died and that, in contrast, adults had similar levels of these same molecules (McElroy *et al*., 2014b).

EBOV initially undergoes massive replication in immune cells: macrophages and dendritic cells (DCs) and then spreads to many other cell types, infecting all vital organs (Zampieri *et al.*, 2007; Falasca *et al.*, 2015). This high viraemia is associated with massive production of pro-inflammatory cytokines and chemokines, lymphocyte apoptosis and the destruction of local tissues. This damage leads to increased endothelial permeability and therefore to multiorgan failure and systemic clotting disorders (Zampieri *et al.*, 2007; Liu *et al.*, 2017). These deleterious effects are heightened in fatal compared to non-fatal or indeed asymptomatic infections (Leroy *et al.*, 2001). These latter cases are thought to be relatively rare, but are obviously challenging to identify.

The hypersecretion of cytokines are most notable in the pro-inflammatory group, including interleukin (IL)-1b, IL-1RA, IL-6, IL-8 and IL-15, along with a number of chemokines and growth factors such as MIP-1a, MIP-1b and IP-10 (Wauquier *et al.*, 2010). Many of these cytokines are also found to be readily produced in non-human primates (NHPs) after challenge with EBOV (Geisbert *et al.*, 2003; Ebihara *et al.*, 2011). The resulting cytokine storm is a key part of the immunopathology caused by EBOV infection.

During EBOV infection, T-lymphocytes in the blood are seen to be depleted, along with natural killer cells (Reed *et al.*, 2004; Wong *et al.*, 2014). The individuals who fail to produce a robust immune response are more likely to suffer a fatal infection. The response to EBOV infection is a careful balance which must be maintained. A sufficient immune response is required, but an uncontrolled response is what causes much of the immunopathology associated with the virus.

Due to the international contribution involved in the 2013-2016 outbreak, analysis of four individuals undergoing treatment for EVD in the USA provided an opportunity to study acute EVD in the absence of outbreak conditions. Here, patients were observed to exhibit a robust proinflammatory response, which supports much of what has been concluded following *in vivo* experiments with NHPs. However, this challenged the potential immune-dampening activities of the virus itself.

1.6 Known determinants of patient outcome

The processes that lead either to survival or a fatal infection are not well known although viral load is known to be a key determinant. This is especially relevant at extremes, where patients with very high viral loads are more likely to suffer a poor prognosis. Factors influencing patient outcome in EVD cases include access to palliative care, anti-viral treatment, age and the presence of co-morbidities.

1.6.1 Palliative care

Supportive care has been suggested to be one of the key factors that helped reduce the average fatality rate across the 2013-2016 EBOV outbreak, which initially sat at over 70% (Lamontagne *et al.*, 2018). The scale of this outbreak meant that the provision of this palliative care was even more challenging than those previous, being impossible in some areas at the peak of the outbreak due to limitations of infrastructure and resources (Fowler *et al.*, 2014). Oral hydration is critical in infections that cause similar symptoms to EVD - i.e. diarrhoea, volume depletion and shock or organ hypoperfusion – such as cholera. In such cases the fatality rate can be greatly reduced upon rehydration of patients (Mahalanabis *et al.*, 1973).

1.6.2 Age

In the 2000-2001 outbreak of SUDV in Uganda, the mortality rate was seen to be greatly affected by age, with paediatric patients seen to experience lower fatality rates (McElroy *et al.*, 2014a). Following analysis of serum samples from this outbreak, the surviving paediatric patients were seen to have elevated levels of certain cytokines and chemokines, which may have played a significant role in their survival. The same

difference was not observed in adults, which suggests that EVD patients of differing ages could require different treatments. Individuals aged 60-64 were at highest risk (McElroy *et al.*, 2014a). Conversely, in the 2013-2016 outbreak, children under five were seen to be at higher risk of death (Cherif *et al.*, 2017).

1.6.3 Coinfections

A large proportion of the patients presenting to the Ebola treatment centre (ETC) tested positive not only for EBOV but also *Plasmodium falciparum*, indicative of an active malaria infection. Malaria as a co-infection has been implicated in the outcome of EVD, with groups finding that it can have beneficial (Rosenke *et al.*, 2016) or deleterious effects (Waxman *et al.*, 2017). Malaria has long been associated with co-morbidity and as a complicating factor in a number of viral infections, from HIV to EBV and human parvovirus B19 (Rénia and Potter, 2006; Toan *et al.*, 2013; Waxman *et al.*, 2017). Guinea experiences high levels of malaria transmission (>1 case per 1000 population), with *P. falciparum* being responsible for 100% of cases in 2013 (WHO, 2014).

Other coinfections have the potential to aid the immune response in battling EVD. The presence of GB virus C (GBV-C) (formally known as hepatitis G virus) was examined in 49 patients positive for EBOV from Sierra Leone, where both survival and GBV-C status were found to be associated with age (Lauck *et al.*, 2014). There have been several case reports of Gram-negative septicaemia in patients with EVD (Kreuels *et al.*, 2014), although very little data exist regarding the frequency of this septicaemia. However, the incidence of bacteraemia appeared to be low upon clinic admittance (Lamb *et al.*, 2015). Despite this, assessing the presence of bacterial infection of the blood from

large numbers of patients infected with high-consequence infections is complicated, particularly under field conditions of high containment, where the primary need is to provide an accurate diagnosis of EBOV. Therefore, bacterial coinfections could frequently be missed.

1.7 RNA sequencing

Transcriptomics, achieved via next generation sequencing (NGS), have been revolutionary in biomedical research. This allows for the abundance of transcripts to be calculated, with a greater abundance suggesting an enhanced gene expression. In RNA sequencing (RNAseq), transcripts are quantified and the relative abundance of transcripts can be elucidated. This allows the activity of genes and their downstream pathways to be investigated. Due to the high containment nature of EBOV, transcriptomics provides an avenue to allow the study of the host response to the virus, as well as sequence changes in the virus itself. NGS is particularly useful in studying the dynamics of viral quasispecies.

1.7.2 Illumina and Nanopore sequencing

Illumina sequencing enables the generation of millions of accurate reads from an initial sample. Reads are roughly 150 bp in length and many known pipelines exist for the study of Illumina data, due to its established nature.

Like Pac-Bio, Nanopore sequencing generates reads of substantial length, up to many thousands of base pairs. Here, RNA binds to pores on a membrane, which passes the nucleic acid through the pore (Figure 1.5). This leads to a change in voltage that is

measured as a “squiggle”, giving a read-out of the predicted base at that position. The major advantage of Nanopore sequencing is that it can be performed rapidly and with limited training, while the error rates produced by this platform are much higher than that of Illumina.

These two techniques are both very useful in the study of viral infections, each with optimal uses and advantages. As such, both have been employed in this thesis.

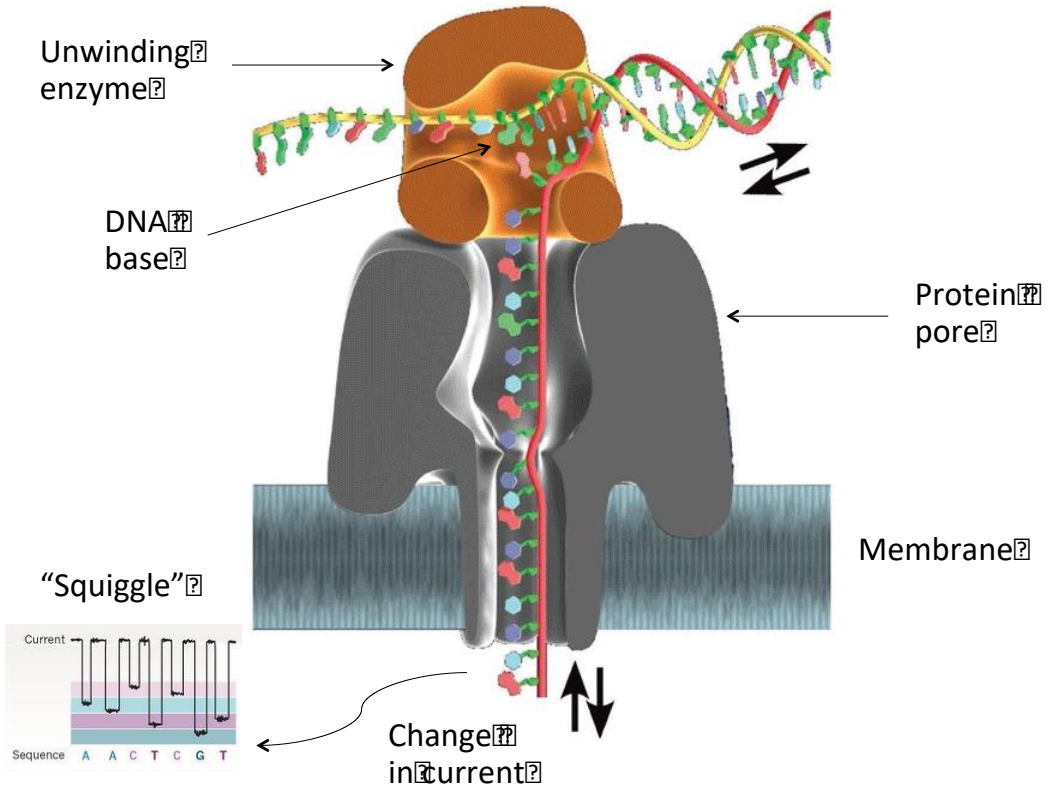


Figure 1.5: Schematic diagram of a Nanopore sequencing DNA. *Here, DNA has been accepted into the nanopore (grey), which is controlled by the unwinding enzyme (brown). The nanopore is embedded in a lipid bilayer and an electric field applied. This electric field is capable of controlling the direction of the DNA through polymerase activity. DNA sequence information is recorded in the form of a "squiggle" – i.e. changes in the ionic current running through the pore. Adapted from Schneider and Dekker (2012).*

1.8 Research objectives

The aim of this project is to use transcriptome data from acutely infected and convalescent patients to identify factors that are associated with patients that survived or succumbed to EVD. This is likely to be a complex analysis and a mixture of viral, host and external factors. The role of the viral sequence itself as well as the potential role of co-infections have not been well investigated previously, while the host response is known to readily affect patient outcome. The 2013-2016 outbreak in West Africa and the involvement of the EMLab has provided a unique opportunity to study this virus and its effects using a comparatively large sample population.

Chapter 2: Materials and methods

All virus preparation and viral infections were performed by collaborators at Public Health England (PHE), Porton Down under CL4 conditions, where I assisted in the downstream processing of samples following viral inactivation. The Illumina sequencing was performed by a team at the Centre of Genomic Research (CGR) at the University of Liverpool and Dr. David Matthews at the University of Bristol.

2.1 Patient samples

2.1.1 Ethical approval

Ethical approval for use of residual diagnostic RNA extracts from blood/other bodily fluids from individuals suffering from acute Ebola virus disease was obtained from the National Committee of Ethics in Medical Research of Guinea (permit number 11/CNERS/14). Informed consent was not obtained from patients due to the nature of the sample collection as part of the public health response to the 2013-2016 outbreak of Ebola.

2.1.2 Human samples

All samples in this study were provided by the Bernhard-Nocht Institute for Tropical Medicine (BNI) in partnership with the European Mobile Laboratory (EMLab). They were taken from EMLab sites in Coyah and Guéckédou (Guinea) during the 2013-2016 Ebola virus outbreak in West Africa. During my PhD I deployed to Coyah as a Team Leader to facilitate diagnostics and retrieval of these samples. The aim of these

laboratories was to provide Ebola diagnostics to the adjacent Ebola Treatment Centre (ETC) and surrounding areas.

At the ETC, 10 ml of blood was taken from a patient suffering from acute symptoms. 50 μl EDTA-whole blood was inactivated in the glove box via the addition of 560 μl of both AVL (Anti-Viral Lysis) and ethanol. Other bodily fluids were accepted (urine, breast milk, semen) for which 140 μl of the original sample was required. Additionally, oral swabs were taken from community deaths which were suspended in solution prior to being treated in the same way as urine. This was done in order to track the spread of the virus.

As this thesis continued, more samples were procured for analysis beyond the ones I processed in Guinea and transported to Liverpool. Therefore, each chapter starts with a table indicating the samples involved.

2.2 Ebola diagnostics

2.2.1 Virus inactivation and RNA extraction

RNA was extracted from EDTA-whole blood using the QIAamp Viral RNA Mini Kit (Qiagen) protocol. Following the addition of ethanol, the sample was vortexed and incubated for 10 minutes. Then, the ethanol was added and mixed again. Once inactivated, the samples were removed from the glove box securely into a bucket of bleach and transferred into the extraction room.

630 µl of this solution was applied to a column in a 2 ml collection tube and centrifuged at 8000 rpm for 1 minute. The filtrate was discarded, and this step repeated. After the second centrifugation, 500 µl Buffer AW1 was added to the column and spun again. This filtrate was discarded and 500 µl of Buffer AW2 added. This was centrifuged at 14 000 rpm for 3 minutes. Following the discarding of the filtrate, 60 µl of Buffer AVE was added to the column and incubated for 1 minute. This was then centrifuged at 8000 rpm for 1 minute and the eluate used for PCR or stored at -20 °C until use. Here the quality and quantity of the RNA was assessed by Nanodrop (Thermo Scientific).

2.2.2 Diagnostic RT-PCR for Zaire ebolavirus

EBOV RNA was detected using the RealStar Zaire Ebolavirus RT-PCR kit version 1.0 (Altona Diagnostics, Hamburg, Germany), using the SmartCycler II system (Cepheid) or Rotor-Gene Q system (Qiagen). These systems produced very similar cycle thresholds (Cts) when compared (Rieger *et al.*, 2016). This kit includes an internal control (IC), which serves as a control of the nucleic acid extraction procedure.

Briefly, 20 µl of Master Mix is added to 10 µl sample or control and mixed thoroughly. This was centrifuged for 30 seconds at 3000 rpm and PCR conditions as in Table 2.1.

Table 2.1: PCR cycling conditions for RealStar Ebolavrius RT-PCR.

	Stage	Cycle repeats	Acquisition	Temperature	Time (minutes)
Reverse transcription	Hold	1	-	55 °C	20:00
Denaturation	Hold	1	-	95 °C	2:00
Amplification	Cycling	45	-	95 °C	0:15
			√	58 °C	0:45
			-	72 °C	0:15

2.2.3 Testing for *Plasmodium spp.* via an immunochromatographic assay

During viral inactivation, 15 μl of the blood sample was isolated for use in a BinaxNow Malaria (Alere) rapid diagnostic test (RDT) for *Plasmodium spp*. This blood sample was transferred onto the sample pad and 2 drops of Reagent A added below. This caused the sample to run up the length of the test strip. At this point, 4 drops of Reagent A was added to the second reagent pad. The RDT was then closed and left for 15 minutes. The interpretation of the results was clear, and a band would appear if the sample was testing positive, along with the control band.

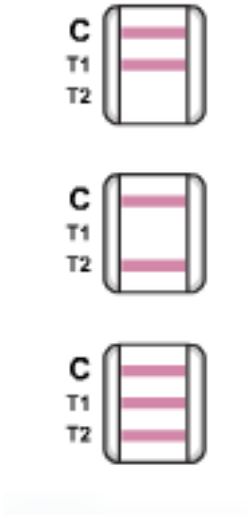


Figure 2.1: Interpretation of Plasmodium spp. rapid diagnostic test (BinaxNOW, Alere). *T1 positive – positive result for P. falciparum; T2 positive – positive result for P. vivax, P. malariae or P. ovale, or a mixture of two or more of those species; T1 + T2 positive – positive result for P. falciparum and a possible mixed infection.*

2.3 qRT-PCR techniques

2.3.1 DNA removal and sample clean-up

TURBO DNase (Ambion, Thermo Scientific) was used for the pre-treatment of samples. Equimolar concentrations of samples were treated with 2U of TURBO DNase for 15 minutes at 37 °C, before heating to 95 °C for 5 minutes and then cooled on ice. For clean-up, the RNEasy MinElute kit (Qiagen) was used. Samples treated with DNase were incubated with 350 μl Buffer RLT (Qiagen) for 10 minutes and then 350 μl 70% molecular grade ethanol added and mixed thoroughly.

This solution was transferred to RNEasy MinElute spin columns cooled at 2-8 °C prior to use and the manufacturer's instructions followed. The resulting purified RNA was eluted in 14 μl nuclease-free water (Qiagen) and concentrations (ng / μl) estimated using a Nanodrop 2000. Samples were diluted to equimolar concentrations prior to downstream activity.

2.3.2 RT2 Profiler PCR array: Inflammatory cytokines and receptors

To profile the abundance of mRNA corresponding to the inflammatory response the RT2 Profiler PCR array: Inflammatory cytokines and receptors (Qiagen, PAHS-011Z) was used. ABI Fast 96-well plates (Life Technologies) were prefilled with lyophilised commercial assays targeting 84 genes associated with the human inflammatory response. The full list can be found in Table 2.2 and Appendix 7.3. Each well contains forward and reverse primers and probes targeting the mRNA of these genes.

For each RNA sample, a genomic DNA elimination mix was prepared. This comprised of 500 μg RNA, 2 μl Buffer GE and enough RNase-free water to bring the total volume up to 10 μl. This mixture was incubated for five minutes at 42 °C and then placed on ice for a minimum of 60 seconds.

A reverse-transcription mix was prepared consisting of 4 μl 5x Buffer BC3, 1 μl Control P2, 2 μl RE3 Reverse Transcriptase Mix and 3 μl RNase-free water. This 10 μl mixture was added to each genomic DNA elimination mix and mixed gently. Following this addition, the mixture was incubated for 15 minutes at 42 °C, then for 5 minutes at 95 °C to stop the reaction. 91 μl RNase-free water was then added to each reaction and mixed thoroughly on ice.

In a 5 ml tube, 1350 μl 2x RT^2 SYBR Green Mastermix was added to the 102 μl cDNA synthesis reaction and 1248 μl RNase-free water. Using a multi-channel pipette, 25 μl added to the PCR array with thorough mixing to help suspend the lyophilized primer mix. This 96-well plate was then sealed and centrifuged for 1 minute at 1000 g at room temperature in order to remove any bubbles before qRT-PCR.

Ct values for all wells were exported to a blank Excel spreadsheet. Data analysis was then conducted at QIAGEN'S GeneGlobe Data Analysis Center using a software-based tool that allows the analysis of RT-PCR data (QIAGEN, 2016). The real-time PCR modules transformed the Ct values to calculated results for gene expression.

Table 2.1: Full list of gene targets tested using a focussed, gene specific profiler array for the human inflammatory response. *Functional categories are shown alongside the accompanying gene target list.*

Chemokines	C5, CCL1 (I-309), CCL11 (eotaxin), CCL13 (MCP-4), CCL15 (MIP-1d), CCL16 (HCC-4), CCL17 (TARC), CCL2 (MCP-1), CCL20 (MIP-3a), CCL22 (MDC), CCL23 (MPIF-1), CCL24 (MPIF-2 , Eotaxin-2, MPIF-2, Eotaxin-2), CCL26 (eotaxin-3), CCL3 (MIP-1A), CCL4 (MIP-1B), CCL5 (RANTES), CCL7 (MCP-3), CCL8 (MCP-2), CX3CL1, CXCL1 (GRO1, GROa, SCYB1), CXCL10 (INP10), CXCL11 (I-TAC, IP-9), CXCL12 (SDF1), CXCL13, CXCL2 (GRO2, GROb, SCYB2), CXCL3, CXCL5 (ENA-78, LIX), CXCL6 (GCP-2), CXCL9 (MIG).
Chemokine Receptors	CCL13 (MCP-4), CCR1, CCR2, CCR3, CCR4, CCR5, CCR6, CCR8, CX3CR1, CXCR1 (IL8RA), CXCR2 (IL8RB).
Interleukins	IL13, IL15, IL16, IL17A, IL17C, IL17F, IL1A, IL1B, IL1RN, IL21, IL27, IL3, IL33, IL5, IL7, CXCL8, IL9.
Interleukin Receptors	IL10RA, IL10RB, IL1R1, IL5RA (CD125), IL9R.
Other Cytokines	AIMP1 (SCYE1), BMP2, CD40LG (TNFSF5), CSF1 (MCSF), CSF2 (GM-CSF), CSF3 (GCSF), FASLG (TNFSF6), IFNA2, IFNG, LTA (TNFB), LTB, MIF, NAMPT, OSM, SPP1, TNF, TNFSF10 (TRAIL), TNFSF11 (RANKL), TNFSF13, TNFSF13B, TNFSF4 (OX40L), VEGFA.
Other Cytokine Receptors	TNFRSF11B (OPG).

2.3.3 Targeted qRT-PCR

Primers were designed by NCBI/Primer-BLAST targeting 2 exons that spans an intron based on the alignments of all available RNA transcript sequences of each of the 10 human genes from the transcriptome database AceView (mRNAs from GenBank or RefSeq, and single pass cDNA sequences from dbEST and Trace). Primers were then synthetized by Eurofin Genomics (Ebersberg, Germany).

The optimised conditions for the one-step singleplex RT-qPCR assay were as follows: 5 µl of RNA added to a reaction mixture of 2X iTaq universal SYBR green reaction mix (2X) and iScript Reverse transcriptase (2X)(Bio-Rad), 100 nM of each primer in a final reaction volume of 20 µl. The one-step singleplex RT-qPCR was run on the CFX 96™ real time PCR machine (Bio-Rad). The following thermal cycling parameters were used: 1 cycle of RT for 10 min at 50°C followed by 1 cycle of polymerase activation and DNA denaturation at 94°C for 1 minute, and 40 cycles of denaturation at 94°C for 10 minutes and annealing/extension at 60°C for 30 minutes.

The measurement of number of RNA copies was determined using a previously developed formula. The formula gives the molecules per µl (N), if the concentration of the cRNA (C) is known in relation to the fragment size (K) multiplied by a factor derived from the molecular mass and the Avogadro constant:

$$N \text{ (molecules per ul)} = \frac{C \text{ (cRNA µg/µl)}}{K \text{ (fragment size/b)}} \; 182.5 \times 10^{13}$$

The efficiency and dynamic range of the respective singleplex assays were established by amplifying five 10-fold dilutions from 10^7 to 10^3 of RNA transcripts of each gene. Each dilution was amplified in triplicate in separate runs by each gene.

2.4 *In vitro* tissue culture techniques

2.4.1 Routine culture of THP-1 cells

THP-1 cells are semi-adherent/suspension cells and were acquired from the ECACC. Here, an authenticated, mycoplasma and bacteria-negative cell bank was established by ECACC technicians. All cultures were started from this authenticated stock to ensure standardisation of experiments across the project. For passaging, cells were counted using plastic disposable C-Chip Neubauer cytometers (NanoEnTek) and reseeded at a density of 0.1-0.3x10e+6 cells per ml. Cultures were maintained between 0.3x10e+6 and 2x10e+6 in RPMI media supplemented with GlutaMAX (Thermo Scientific) at 10% for routine culture.

2.4.2 Generation of PMA-differentiated THP-1 macrophage-like cells

Naïve THP-1 cells are semi-adherent and morphologically and functionally similar to monocytes and can be differentiated into macrophage like cells using Phorbol-12-myristate (PMA, Sigma-Aldrich) (Qin, 2012; Genin *et al.*, 2015). PMA was diluted in DMSO to a working concentration of 1 μg/ml, then to a final concentration of 5 ng/ml in RPMI supplemented with 10% media. Cells were suspended in complete growth media with additive 5 mg/ml PMA to an appropriate concentration. After thorough mixing by inversion, cells were added to flasks to form 80% to 90% confluent layer of cells. The growth of PMA differentiated THP-1 cells will be inhibited, therefore

activation was initiated after the required cell volume was reached. 24 hours after the addition of PMA, cells were rinsed twice with PBS (Thermo Scientific) and fresh complete RPMI growth media added. Cells were rested for 72 hours before infection took place, during which time the cells changed in their morphology, becoming adherent.

Cell culture in CL4 conditions provided challenges in the form of high levels of formaldehyde residues and cytotoxic chemicals. The result is that cell growth was maintained in flasks with plug-sealed caps and locked within sealable boxes. In negative controls, cell viability remained high (> 90%), as evaluated using Trypan Blue (Sigma-Aldrich) and manual counting using a cytometer.

2.4.3 Viral infection

The EBOV Makona variant was obtained from the European Mobile Laboratory and was isolated during the laboratory response to the 2013-2016 West African outbreak. This isolate was designated Ebola virus/H. sapiens-wt/GIN/2014/Makona-Gueckedou-C05. This isolate has been cultured in Vero cells. Virus was quantified by a modified $TCID_{50}$ method.

From these infected cells, viral RNA was extracted as described in 2.2.1 and quantified as described in 2.2.2.

2.5 Nanopore sequencing

2.5.1 RNA preparation

The cDNA-PCR Sequencing Kit (SQK-PCS108) was used to prepare samples for MinION (Oxford Nanopore) sequencing. The samples were RNA extracts from retrieved from the EMLab as described previously. RNA extracts were quantified using a Qubit (Thermo Scientific) where the Broad Range RNA kit (Abi, Thermo Scientific) was used. A Nanodrop 2000 (Thermo Scientific) assay for RNA concentration estimation was performed and ratios at 260/280 nm and 260/230 nm calculated to determine sample purity. 50 ng was transferred into a DNA LoBind tube and the volume adjusted up to 9 μl with nuclease-free water. This was mixed via flicking of the tube and spun down using a microfuge.

2.5.2 Reverse transcription and strand-switching

To the previous solution, 1 μl VNP and 1 μl 10mM dNTPs were added and mixed gently. This solution was incubated at 65 °C for five minutes and then cooled. In a separate tube, a master mix was composed consisting of 4 μl Superscript IV buffer, 1 μl RNaseOUT, 1 μl 100 mM DTT and 2 μl Strand-Switching Primer (SSP). This mixture was added to the annealed mRNA and mixed thoroughly. This mixture was incubated at 42 °C for two minutes and then 1 μl of SuperScript IV Reverse Transcriptase was added to bring the total volume for each sample up to 20 μl. This was mixed gently and spun down before being incubated following the conditions in Table 2.3.

Table 2.3: Heating conditions for reverse transcription.

Function	**Length (minutes)**	**Temperature (°C)**	**Number of cycles**
Reverse transcription	10	50	1
Strand switching	10	42	1
Heat inactivation	10	80	1
Hold	-	4	-

Within the reaction tube, 25 μl 2x LongAmp Taq Master Mix, 1.5 μl cDNA PRM (cPRM) and 18.5 μl nuclease-free water were added to 5 μl reverse-transcribed RNA sample. Four PCR reactions are carried out for each RNA sample and then pooled. This solution was involved in an amplification step following the following conditions (Table 2.4).

Table 2.4: Reverse transcription conditions using a CFX96 (Bio-Rad). ** In order to maximise read length an extension time of two minutes and 30 seconds was used.*

Function	**Length (seconds)**	**Temperature (°C)**	**Number of cycles**
Initial denaturation	30	95	1
Denaturation	15	95	18
Annealing	15	62	18
Extension	50/kb*	65	18
Final extension	6 minutes	65	1
Hold	-	4	-

The four PCR reactions for each sample were pooled and added to 160 μl AMPure XP beads before a five minute incubation on a rotator mixer at room temperature (RT). The samples were spun down and pelleted using a magnetic rack, allowing the

supernatant to be removed. The beads were washed using 70% ethanol twice. After the 70% ethanol was removed for the second time, the beads were allowed to dry briefly. The beads were then resuspended in 21 μl Rapid Annealing Buffer (RAB) and incubated on the rotator mixer for 10 minutes at RT. After this, the beads were pelleted again on the magnetic rack and this time the eluate removed and retained.

2.5.3 Adapter addition and AMPure clean up

1 μl of this amplified DNA was quantified using the Qubit. 500 ng of cDNA was taken and made up to 20 μl with RAB. To this, 5 μl of cDNA Adapter Mix was added, and this solution spun down. This was followed by a five-minute incubation on a rotator mixer at RT. 20 μl of resuspended AMPure XP beads were then added and the solution allowed to pellet on a magnetic rack. The supernatant was removed and 140 μl ABB buffer added, allowing the beads to resuspend. This process was repeated once more. After pelleting for the second time, the beads were resuspended in 12 μl Elution Buffer (ELB) and incubated using the rotator mixer for 10 minutes at RT. Following this, the beads were pelleted again and 12 μl of eluate removed and retained. This solution was stored on ice until sequencing.

2.5.4 Priming and loading of flow cells

In order to prime the flow cells for use, a very small volume of liquid was removed from the priming port, in order to determine that no bubbles were present. 576 μl RBF was then added to 624 μl nuclease-free water to form a priming mix. 800 μl of this was loaded into the flow cell via the priming port.

In order to load the library, 12 μl DNA library was added to 35 μl RBF, 2.5 μl nuclease-free water and 25.5 μl LLB. The remaining 200 μl of the priming mix was loaded via the SpotON sample port followed by 75 μl of the sample in a drop by drop delivery method. Here, care must be taken in order to prevent the introduction of any bubbles, as the effect of this could be substantial. After this, the sequencing run can begin using MinKNOW software (Oxford Nanopore).

2.6 Next generation sequencing

2.6.1 DNA removal and clean-up

RNA samples that went on to be Sequenced on a HiSeq 2500 or 4000 (Illumina) were first treated with TURBO DNase (Ambion, Thermo Scientific). Samples were treated with 2U of TURBO DNase for 15 minutes at 37 °C, before heating to 95 °C for 5 minutes and cooled on ice. Subsequently, the RNEasy MinElute kit (Qiagen) was used. DNase-treated samples were added to 350 μl of Buffer RLT (Qiagen) and left for 10 minutes. Then, 350 μl 70% molecular grade ethanol was added and the sample mixed.

These samples were transferred to RNEasy MinElute spin columns (pre-cooled to 2-8 °C) and the protocol followed. Purified RNA was eluted in 14 μl Nuclease Free Water (Qiagen).

2.6.2 Illumina library preparation

This was performed by colleagues at the Centre of Genomic Research. Illumina libraries were prepared using the DNase-treated RNA using the Epicentre ScriptSeq v2

RNA-seq library preparation kit and performing 10 to 15 cycles of amplification. Libraries were purified using AMPure XP magnetic beads (Ambion). Each library was quantified using Qubit (Thermo Scientific) using the DNA High Sensitivity Kit, the size distribution was assessed using the Agilent 2100 Bioanalyser, and the final libraries were pooled in equimolar ratios. The quantity and quality of each pool were assessed with the Bioanalyzer and subsequently by qPCR using the Illumina library quantification kit (Kapa) on a Roche Light Cycler LC480II system according to the manufacturer's instructions.

2.6.3 HiSeq 2500 loading

The template DNA was denatured according to the protocol described in the Illumina User Guide and loaded at 12 pM. To improve sequencing quality control, samples were spiked with 1% PhiX, functioning as an IC viral RNA. The sequencing was undertaken on the Illumina HiSeq 2500 with version 4 chemistry, generating 2 x 125-bp paired-end reads. Base calling and demultiplexing of indexed reads were performed by using Casava version 1.8.2 (Illumina) to produce all the sequence data in fastq format.

2.6.4 Sequencing of guinea pig samples

Animal experiments were performed as described in Dowall *et al.* (2014). Pooled spleens of six infected guinea pigs were treated as above but the pool of libraries was sequenced on one flow cell of the MiSeq at 2 × 150 based paired-end sequencing with v2 chemistry.

2.7 Bioinformatics analysis

2.7.1 Next generation sequencing data analysis in GALAXY (Penn State)

GALAXY is an online system that provides a user-friendly interface, accessing a library of command line scripted algorithms for the analysis of next generation sequencing (NGS) data.

The raw fastq files were trimmed to remove Illumina adapter sequences by using Cutadapt version 1.2.1. The option '-O 3' was set, so the 3' end of any reads which matched the adapter sequence over at least 3 bp was trimmed off. The reads were further trimmed to remove low-quality bases, using Sickle version 1.200 with a minimum window quality score of 20. After trimming, reads shorter than 10 bp were removed. If both reads from a pair passed this filter, each was included in the R1 (forward reads) or R2 (reverse reads) file. If only one of a read pair passed this filter, it was included in the R0 (unpaired reads) file.

2.7.2 Mapping of reads and fusion events

The resulting R1/R2 sequence read pairs were initially mapped to the human reference genome GRCh38 (ftp://ftp.ensem-bl.org/pub/release-77/fasta/homo_sapiens/dna/Homo_-apiens.GRCh38.dna_sm.primary_assembly.fa.gz) using Tophat2 version 2.1.0 using paired-end mapping, which used the short-read mapper Bowtie2 (Langmead and Salzberg, 2012). Those reads that did not map to the human genome were then pooled and mapped again to a list of known human transcripts again using Bowtie2. Read counts were generated from these alignments

using HTSeq-count and a gtf file, which informed where the gene boundaries sit on the reference genome. In order to detect fusion events in the EBOV genome samples were mapped to EBOV Makona-Gueckedou-C05 using Tophat2 and the fusion search enabled. The anchor length was set to 20, meaning that a supporting read had to map both sides of a fusion event by 20 bases. Reads containing fusion events that mapped to multiple parts of the genome were excluded (Kim and Salzberg, 2011).

2.7.3 Identification of differentially expressed genes using EdgeR

EdgeR calculates fold change in a transcript comparing hospitalised fatalities to hospitalised survivors and that acutely ill cohort to convalescent controls. The normalised read counts generated by HTSeq-count were imported into R. EdgeR uses negative binomial distribution in order to model read counts and provide a p value as well as a false discovery rate (FDR), and therefore the adjusted statistical significance of supposed differences (Robinson *et al.*, 2010). An absolute log2(fold change) > 1 was considered to be differentially expressed, with an FDR < 5%.

2.7.4 Identification of microorganism transcripts

This was performed by colleagues at the Centre of Genomic Research. Sequences that did not align to either the human genome or human transcriptome were pooled, but the size of the data set (~1 TB of fastq reads) required them to be pre-processed first by the Trinity read normalization software to reduce the memory requirements and runtimes prior to de novo assembly using Trinity (Haas *et al.*, 2013). The assembled data from Trinity were first manually checked to determine that the EBOV genome had been assembled before all the reads were checked for matches using Diamond BLASTX analysis against the nonredundant protein database (Buchfink *et al.*, 2015). For each

transcript, the best hit was retained and the protein name and organism for that hit was associated with the transcript using in-house scripts. Quality and primer trimmed reads were used as input for MetaPhlAn 2 for estimating genome coverage of bacterial species for each sample. In general, for viral species a 20% mapping to a transcriptome was adopted as a threshold for positive identification.

2.7.5 KEGG analysis

The Kyoto Encyclopedia of Genes and Genomes (KEGG) database consist of proteins, genes and their functional properties (Kanehisa and Goto, 2000). Therefore, canonical pathways can be highlighted once linked to a dataset via determining which pathways are the best fit using the genes that are most represented in any one pathway. When this tool was employed the default settings were used and the *Homo sapiens* (HSA) pathway database selected and an adjusted p value of 0.05 chosen as cut-off.

2.7.6 Quasispecies analysis

QuasiBAM, developed by Public Health England, was used to measure the genome diversity in the generated bam files. This tool functions by comparing the assembled, mapped reads against a reference genome and identifying the minor and major variants, i.e. those nucleotide changes occurring at a high and low frequency respectively.

2.7.7 Nanopore data analysis using GALAXY (Penn State)

MinKNOW outputs sequences into fastq and fast5 files. Using the fastq files, reads were concatenated as single-end reads are generated and dumped when the number of reads reaches 4000. These were trimmed using Porechop v0.2.3, in order to cope

with the “noisy” data generated using Nanopore (Wick *et al.*, 2017), and mapped to the EBOV genome using BWA-MEM, for which Oxford Nanopore has created an analysis mode which fits well with the type of data generated by this platform – Nanopore 2D-reads mode (C. Jain *et al.*, 2018). These files were removed and subsequently mapped to the human genome. These generated bam files were aligned to gtf files - Homo sapiens GRCh38.77 or Zaire ebolavirus ncbiGene. The transcription profiles of hospitalised survivors and hospitalised fatalities were compared using DESeq2.

2.7.8 Ingenuity Pathway Analysis

Ingenuity Pathway Analysis (IPA; Qiagen) is a toolkit available for the analysis of gene and protein expression data. Fold change data is uploaded and annotated in the software. The software analyses for over-represented canonical pathways, functional groups and disease groups. Additionally, activation score (z-score) is calculated to determine the level of pathway activation using the cumulative fold increase or decrease of the genes aligned to each pathway. This allows the host response to be analysed and patient groups compared.

2.7.9 Digital cell quantification

In order to determine predicted immune cell profiles using transcriptomic data, the R package ComICS was used (Altboum *et al.*, 2014). The default parameters were used with 500 repeats and a split ratio of 50% on an input dataset of fold changes compared to convalescent controls (hospitalised_fatalities to controls and hospitalised_survivors to controls). All 207 cell types were analysed. The resulting data provided a predicted mean relative abundance as well as the standard deviation of the mean. To determine

cell types that were significantly different from zero, a standard t-test was performed with a p value cutoff of 0.05. Definition of immune subsets is based on ImmGen database information and thus represents previously validated mouse equivalents to human immune cells (Auffray *et al.*, 2009; Villadangos and Shortman, 2010; Haniffa *et al.*, 2012).

2.8 Statistical analysis

Statistical analyses were performed in R, SPSS 25 and GraphPad Prism version 6.0 and described here.

2.8.1 General linear model (GLM) and analysis of variance (ANOVA)

As part of the multi-variate analysis, a GLM was generated in order to determine which variables could be potentially influencing patient outcome. This allows multiple models of regression to be run simultaneously in order to determine data patterns and potential correlates of prognosis.

An ANOVA allows multiple groups and/or variables to be compared simultaneously and their variances compared. Multiple comparisons are corrected for using post-hoc analyses. This allowed the testing of the potential effects variables exhibit on datasets.

2.8.2 Principle component analysis (PCA)

PCA uses an algebraic orthognanol transformation to covert variables of potentially correlating data into linearly uncorrelated values or principle components. The number of variables is reduced allowing entire datasets to be described in terms of relatedness

between populations. The first component analysis is the most sensitive and each one thereafter requires more transformation and is therefore less sensitive. By only using two or three components, very complex data can be simplified, compared and separated.

2.8.3 Linear discriminant analysis (LDA)

An LDA uses Fisher's linear discriminant to determine linear combinations of features that can separate variables, similarly to an ANOVA. However, an LDA is able to separate categorical dependent variables – in this case patient outcome – according to continuous independent variables. In order to visualise this analysis, bi-plots and partition plots were generated using the 'MASS' package in R.

Chapter 3: Viral quasispecies and the impact of genome variation on patient outcome during Ebola Virus Disease

3. 1 Introduction

3.1.1 2013-2016 outbreak of EBOV-Makona

The 2013-2016 EVD outbreak in West Africa was unprecedented in scale, allowing an in-depth analysis of viral adaptation within naturally-infected humans and the observation of previously unknown aspects of EBOV biology. The variant of EBOV from this outbreak was given the name EBOV-Makona, after a river running through Guinea, Sierra Leone and along the border of Liberia, the three most-affected countries (Kuhn *et al.*, 2014). High-resolution genome analysis was used to analyse consensus EBOV sequence and nucleotide variation within (intra) and between (inter) individuals (Gire *et al.*, 2014; Carroll *et al.*, 2015; Park *et al.*, 2015). Fitness, as measured by the ability of a virus to replicate with high fidelity and adapt to multiple host pressures, plays a major role in determining viral load and the definition of each patient's disease course.

Investigations into Makona compared to other EBOV isolates have identified few differences in cell entry (Hofmann-Winkler *et al.*, 2015), immune response (Bosworth *et al.*, 2017) and pathogenicity (Marzi *et al.*, 2018), however, the latter is disputed

(Wong *et al.*, 2016; Bosworth *et al.*, 2017). These potential differences highlight the importance of the thorough study of this isolate.

3.1.2 EBOV replication leads to the generation of viral quasispecies

EBOV RNA genome replication fidelity is determined by the inherent error rate induced by the viral RNA-dependent RNA polymerase (the L protein). Additionally, there exist potential genome modifications resulting from errors induced by cellular proteins involved in RNA processing, including Adenosine Deaminases Acting on RNA (ADARs) (inducing an A to G transition) and APOBEC (C to U transition) (Martínez and Melero, 2002; Sheehy *et al.*, 2002; Suspene *et al.*, 2011). Any perturbation in the fidelity of genome replication that increases the rate of mutation can lead to error catastrophe where synonymous and non-synonymous changes lead to a loss in viral RNA synthesis and viral protein function. This can be exploited therapeutically where drugs such as ribavirin can be used to drive the replication of RNA viruses towards being less fit (Crotty *et al.*, 2001; Severson *et al.*, 2002; Graci and Cameron, 2006; Aljabr *et al.*, 2016). Such errors lead to the generation of viral quasispecies. This concept refers to the viral population existing as a consensus viral genome surrounded by other genomes of fluctuating nucleotide variation (Chapter 1, Figure 1.4) (Domingo *et al.*, 2012). Whilst the generation of quasispecies can be exploited therapeutically, it also has benefits for the virus as it allows the successful navigation of population bottlenecks, which can be induced by zoonotic transmission, the host immune response or antiviral treatment.

3.1.3 Persistence of infection

The 2013-2016 outbreak has shown that certain individuals experience persistent infections following EVD. A number of patients have remained EBOV-positive for extended periods in certain body fluids after testing negative in their blood via qRT-PCR. During routine diagnostic testing for the European Mobile Laboratory (EMLab), a study was launched that followed male survivors into their convalescence, testing their semen, urine and saliva for EBOV RNA (Sissoko *et al.*, 2017). During this study, the clearance time for EBOV in seminal fluid was seen to be as long as 13 months, with the first confirmed case of sexual transmission of the virus being recorded (Christie *et al.*, 2015). This is unlikely to be the first time this has occurred, simply the first time it was able to be established with confirmed laboratory diagnosis and genetic analysis. For example, EBOV was found in the semen of an individual following a laboratory accident at Porton Down in 1976 (Emond *et al.*, 1977). Additional cases show viable EBOV in the aqueous humor of the eye three months after viral clearance (Varkey *et al.*, 2015).

There are several hypotheses as to how this persistence is maintained but it is not well understood. One such hypothesis is the formation of defective interfering (DI) genomes. These are truncated viral genomes capable of interrupting normal viral replication via the competition for viral resources, hence resulting in a reduction of virus infectivity (as reviewed by Marriott and Dimmock, 2010). These shorter genomes contain intact 3’ and 5’ ends (and replication signals) but include substantial deletions of coding (and non-coding) sequence. Most commonly identified in *in vitro* experiments, DIs have been detected in a number of viruses including EBOV (Calain *et al.*, 1999) and measles (as reviewed by Rima and Duprex, 2005). Although less

common, DIs have also been identified in naturally-infected humans, for example in patients with acute Dengue (Li *et al.*, 2011) and chronic Hepatitis C (Noppornpanth *et al.*, 2007).

The production of DIs can be compared to the modulation of viral gene expression observed in HIV and HPV among others (Carl *et al.*, 2002; Lace *et al.*, 2008). This modulation also has the ability to produce persistent infections. Immune modulation may also be involved in maintaining these infections, as suggested in Hepatitis B virus (HBV) and HIV-1, among many others (Stoop *et al.*, 2005; Rustagi and Gale, 2014).

3.1.4 RNA-sequencing techniques

Illumina RNA-sequencing (RNA-seq) has been used successfully to study viral populations for many years. However, the technique is not infallible. The generation of short reads makes the study of certain aspects of EBOV biology more challenging. Nanopore sequencing, on the other hand, produces long reads many thousands of nucleotides in length. This has the potential to encompass an entire viral genome in one read and has been shown to be effective in viral sequencing (Quick *et al.*, 2016; Kafetzopoulou *et al.*, 2018). Here, long reads could allow for the investigation of viral quasispecies and DIs with more confidence than using reads generated by an Illumina platform.

3.1.5 Objectives and hypotheses

The aims of this chapter are to understand the potential factors influencing patient outcome of EVD-Makona and identify certain risk groups. RNA extracts have been obtained from the EMLab in Guinea and categorised according to a number of factors

including their age and sex. Also, the minor variants observed along the EBOV genome have been considered and therefore the generation of viral quasispecies. It is possible that these populations will differ between hospitalised fatalities and survivors. This could be in terms of specific mutations associated with fatalities and therefore an increase in pathogenicity, or a more general trend at the viral population level in terms of variation. The 2013-2016 outbreak in West Africa has highlighted the ability of the virus to persist within certain body compartments of infected individuals for extended periods of time (Varkey *et al.*, 2015; Sissoko *et al.*, 2017), with one explanation being DIs. These are predicted to be found in experimentally infected cells, and if found in patient samples, could offer explanation as to how this persistence occurs. The sequencing of RNA extracts from infected cell lines and diagnostic samples has also provided an opportunity to compare Illumina and Nanopore RNA-sequencing. Each has their advantages, with different uses in particular settings.

3. 2 Results

3.2.1 Viral load can be used as a predictor of patient outcome

To identify changes in the viral genetics of EBOV during infection, blood samples from infected patients were analysed via RNA-seq. These samples were collected by the EMLab and myself in Guinea during 2014 and 2015. The samples were taken with the foremost aim of diagnosing the presence of EBOV using quantitative (q)RT-PCR, which was then used in patient management in the Ebola Treatment Centre (ETC). For this purpose, RNA was extracted in the setting of the EMLab in Guinea. Blood samples were taken from acutely ill patients, as well as oral swabs from deceased individuals in

the community. Discarded samples were then held in an archive and used in this study under the auspices of the EVIDENT project– Ebola Virus Disease correlates of protection, determinants of outcome and clinical management. These discarded samples were then analysed by Illumina RNA-seq to identify and quantify viral RNA. Following sequencing of 138 individual samples from individual patients, strict selection criteria were employed to identify and remove datasets from samples that showed evidence of having degraded RNA (expected from field sample collection) (Table 3.1). From an initial set of 138 individual sequenced samples from separate patients, application of these selection criteria led us to discard 26 and analyse the data from 112 unique patients: hospitalised survivors (n = 24) and hospitalised fatalities (n = 88). In the context of this thesis this is defined as follows: blood was collected from hospitalised individuals at a time when the outcome for these patients was unknown and subsequently recorded as they either succumbed to a fatal infection or survived EVD. As observed in previous outbreaks, the viral load of individuals who went on to die from EVD was higher than in survivors (Figure 3.1).

Table 3.1: Summary of patients involved in this study. *These have been split according to sample type and patient outcome. Here, patient age and EBOV Ct are described by the range and mean. N describes the total number of samples in each patient group.*

Patient group		**Hospitalised fatalities**	**Hospitalised survivors**	**Convalescent controls**	**Community deaths**
N		88	24	16	24
% Male		46.3	31.0	93.8	29.1
Age (years)	Range	2 mo – 78	10 mo – 68	18 – 40	1 - 70
	Mean	29.5	32.1	33.1	30.0
EBOV Ct	Range	12.1 – 26.6	15.8 – 31.6	NA	12.4 – 31.9
	Mean	16.9	21.7	NA	18.8

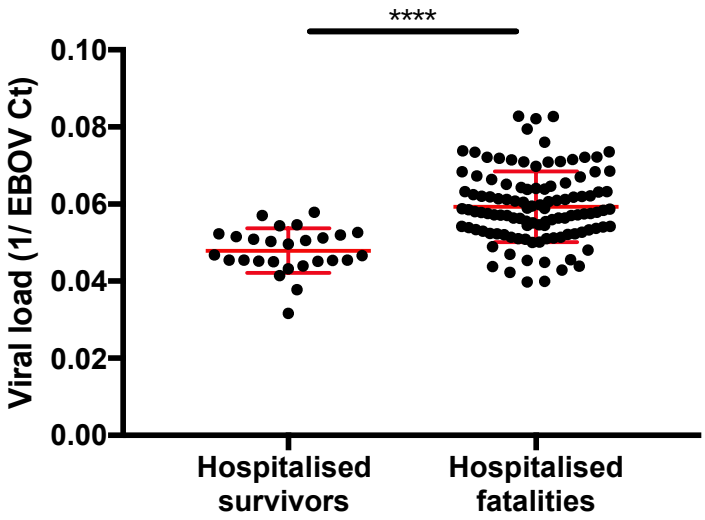


Figure 3.1: EBOV viral load is significantly higher in individuals who die from EVD, compared to those who survive. *Here, the reciprocal EBOV cycle threshold (Ct) value (a proxy for viral load) is compared between acute sufferers of EVD who go on to survive or die. Mann-Whitney U test performed using Prism 7 GraphPad (CA, USA) to compare groups ($p < 0.0001$).*

Importantly, there was no significant difference in the time between symptom onset (as reported by the patient) and taking of the sample during acute illness. The mean time to the onset of symptoms for the acute survivors was 6.4 days and for acute fatalities was 5.9 days with a range of 2–19 and 2–22 days, respectively (Figure 3.2). Here, time between symptom onset and sample procurement was used as a proxy for length of infection. This is an important factor to consider when comparing groups of infected patients. A prediction could be made that the longer an individual has been infected, the more viral replication has occurred, thus widening the viral quasispecies and increasing genome variation. Certain patients were not clinically fit to be providing this information, whilst others claimed to have had acute symptoms for many years. Therefore, strict inclusion criteria were established, with individuals claiming to be symptomatic for more than 30 days not included in the analysis as this is likely to be due to co-morbidities. This cut-off was informed by the incubation period of EBOV and necessary due to the outlying patients – i.e. one individual claimed to be suffering from related symptoms for 256 days.

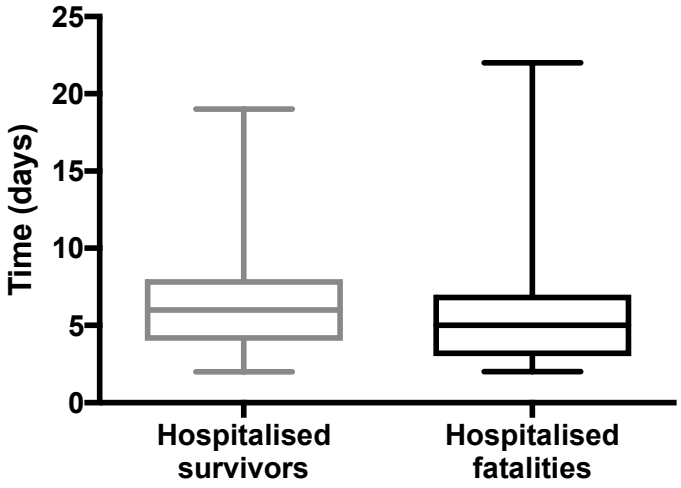


Figure 3.2: Patient groups did not differ in terms of time between symptoms onset and sample procurement. *Mann-Whitney test was performed and no significant difference was found between the groups ($p > 0.1$).*

3.2.2 Potential factors influencing patient outcome

Other criteria were present on the admission form to the ETC, including EBOV Ct value, patient sex, age, response to malaria rapid diagnostic test (RDT) and their geographic location. Analysing this information together allowed the construction of a general linear model (GLM) in order to assess which factors, if any, were found to influence patient outcome. Within these samples, age appeared to have no impact on patient survival (Figure 3.3A, $p > 0.1$), along with sex (Figure 3.3C, $p > 0.1$). Despite this not being statistically significant, a higher proportion of women went on to die from EVD. As mentioned previously, EBOV Ct was significantly lower (demonstrating a higher viral load) in fatal patients (Figure 3.2), as well as those suffering with acute malaria as a co-infection (Figure 3B, $p = 0.053$).

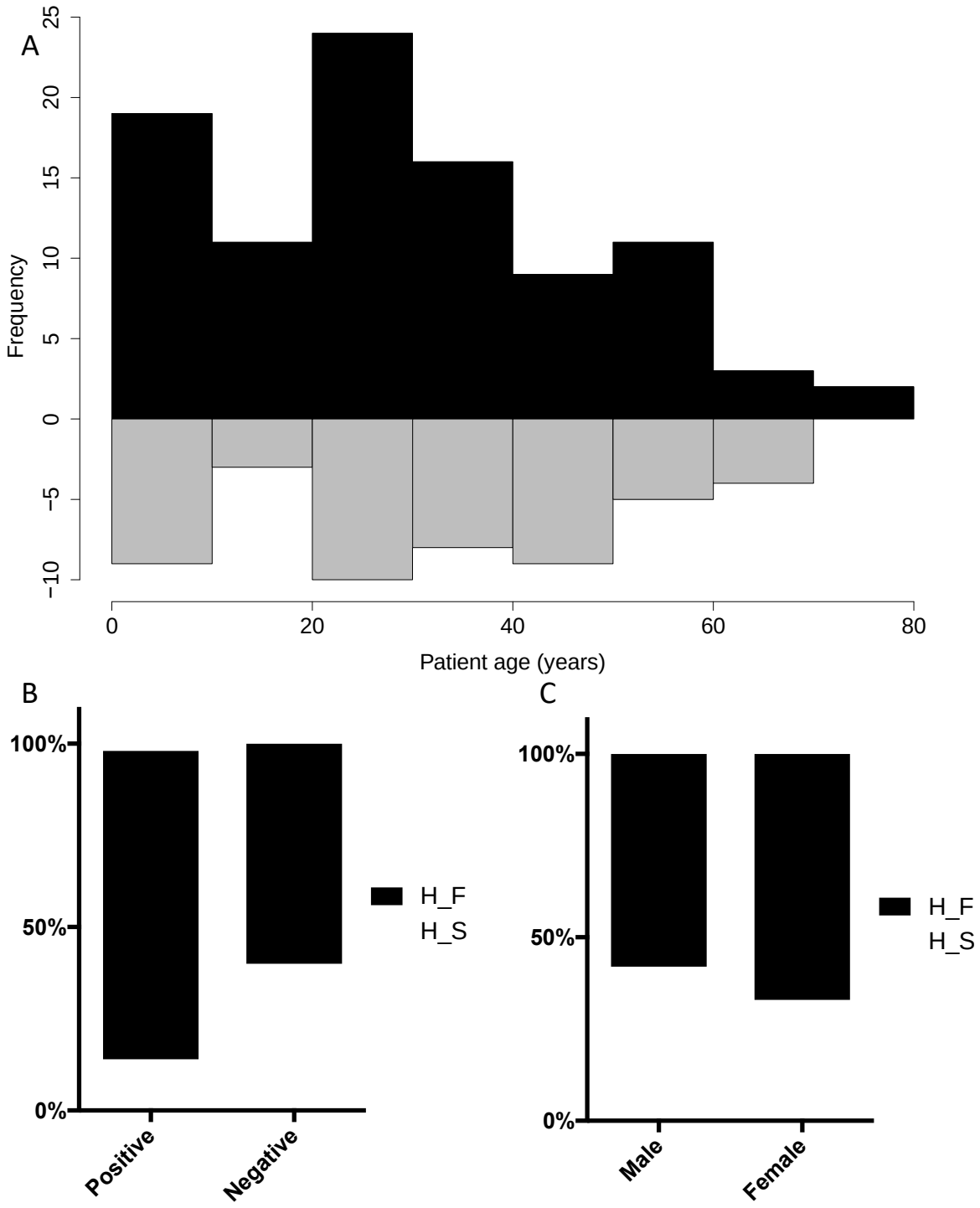


Figure 3.3: Potential factors influencing patient outcome. *Patients split by outcome into hospitalised fatalities (black) and survivors (grey). General linear model (GLM) constructed in order to compare distribution of age (A, p > 0.1), response to Plasmodium spp. RDT (B, p = 0.053) and sex (C, p > 0.1) between patient groups. ANOVA performed using SPSS.*

3.2.7 Comparison of RNA-sequencing techniques

During the outbreak and my deployment with EMLab, Nanopore sequencing was employed for the determination of phylogenetic analysis. This was achieved via the use of EBOV-specific amplicons (Quick *et al.*, 2016). This was also performed immediately following RNA extraction, ensuring the optimal quality of the RNA, with no potential degradation due to freeze-thaw cycles.

In order to determine the further uses for this technique, a transcriptomic approach was employed and three samples from hospitalised survivors were compared to three from hospitalised fatalities. In order to gain a transcriptomic profile, a PCR cDNA sequencing kit was used, employing random primers. This would allow several key questions to be answered. For example, (1) can Nanopore sequencing be employed to build a consensus EBOV genome using clinical samples? (2) Can long reads from Nanopore sequencing elucidate our understanding of viral quasispecies? (3) Can long reads truly confirm the presence of DIs as opposed to simply identifying fusion events?

Upon mapping to the EBOV genome, very few samples had substantial coverage. In order to attempt to build a consensus genome, all six samples were concatenated and mapped again to the EBOV genome, but this still did not provide sufficient coverage. Reads following Nanopore sequencing are much longer than Illumina reads, up to several thousand base pairs in length. There are very few reads mapping to the 5’ end of the genome, with coverage for the L protein being almost non-existent (Figure 3.4A). Therefore, constructing a consensus genome proved impossible.

Comparing a sample sequenced using Nanopore to an Illumina-sequenced run of a sample illustrates the extent of the drawbacks of using Nanopore sequencing to gain information using partially degraded diagnostic samples (Figure 3.4B). These samples are both RNA extracts from diagnostic blood samples that tested positive for EBOV by qRT-PCR. Therefore, the further investigation into understanding EBOV was performed using samples sequenced on an Illumina platform.

Figure 3.4: Coverage across the EBOV genome following Nanopore sequencing. *All reads mapping to EBOV following the sequencing of RNA samples from the extracted RNA of the blood of EVD patients (A). This was performed using the MinION. The coverage of using Nanopore sequencing (blue line) is much lower when compared to a run using Illumina sequencing (black line) (B).*

3.2.3 Differences in viral variation between hospitalised fatalities and hospitalised survivors

Investigating the viral genome diversity in patients appeared to reveal some fundamental differences between patients with the alternate outcomes. Viral diversity was calculated for the patient samples by measuring the minor variants at a threshold of 0.05 across the viral genome using a sliding scale of 100 nucleotides. It is likely that many factors would be capable of influencing the viral diversity found within a patient, including viral load, a patient's stage of infection and the host immune response. In the sample set analysed, there was no correlation between viral diversity and length of infection (Figure 3.5).

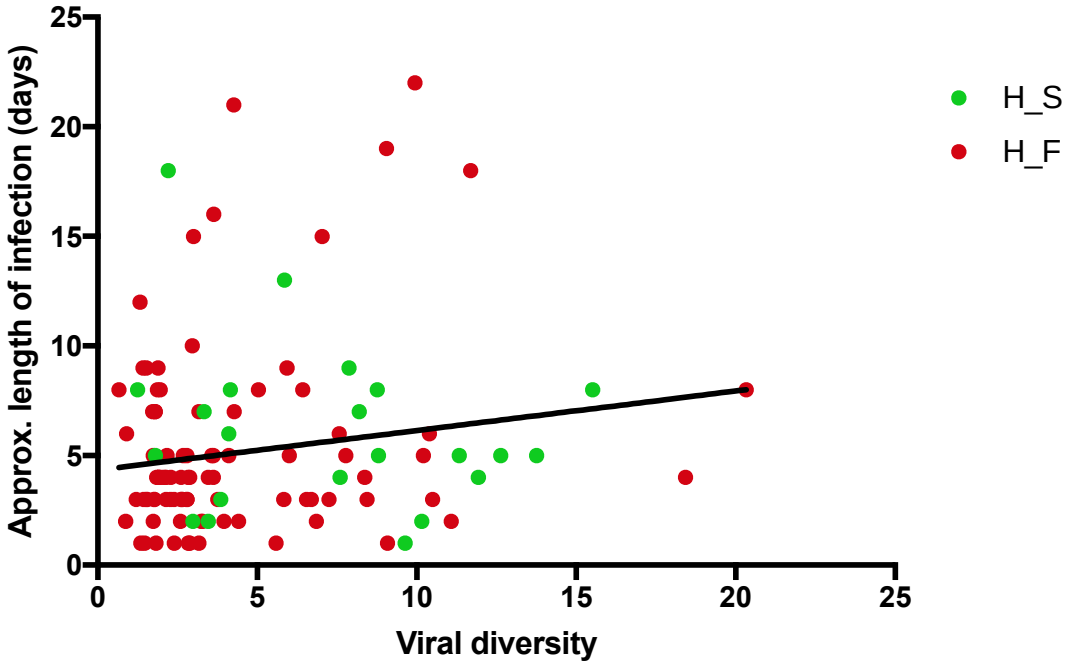


Figure 3.5: Viral diversity did not correlate with an individual's length of infection. *Mean viral diversity was plotted against length of infection, with no significant correlation being observed between the two factors. Linear regression performed (black line), $R^2 = 0.025$, $p = 0.065$. Points are colour-coded by patient group, where H_F are hospitalised fatalities (red) and H_S are hospitalised survivors (green). Viral diversity is a measure of the number of minor variants (at a rate of 0.05 or above) using a sliding window of 100 nucleotides along the length of the EBOV genome.*

Perhaps surprisingly, as viral load decreased, viral diversity was observed to be higher (Figure 3.6A, $p < 0.0001$). Following on from this, patient groups were compared, where the viral diversity in patients who went on to survive infection was significantly higher than those who die (Figure 3.6B, $p = 0.0019$). The samples of a small number of community deaths record the highest genome diversity when compared to the samples from hospitalised individuals. However, this did not reach a level of statistical significance (Figure 3.6B). Any findings drawn from this comparison must be considered carefully due to the fact that these patient groups represent different biological samples.

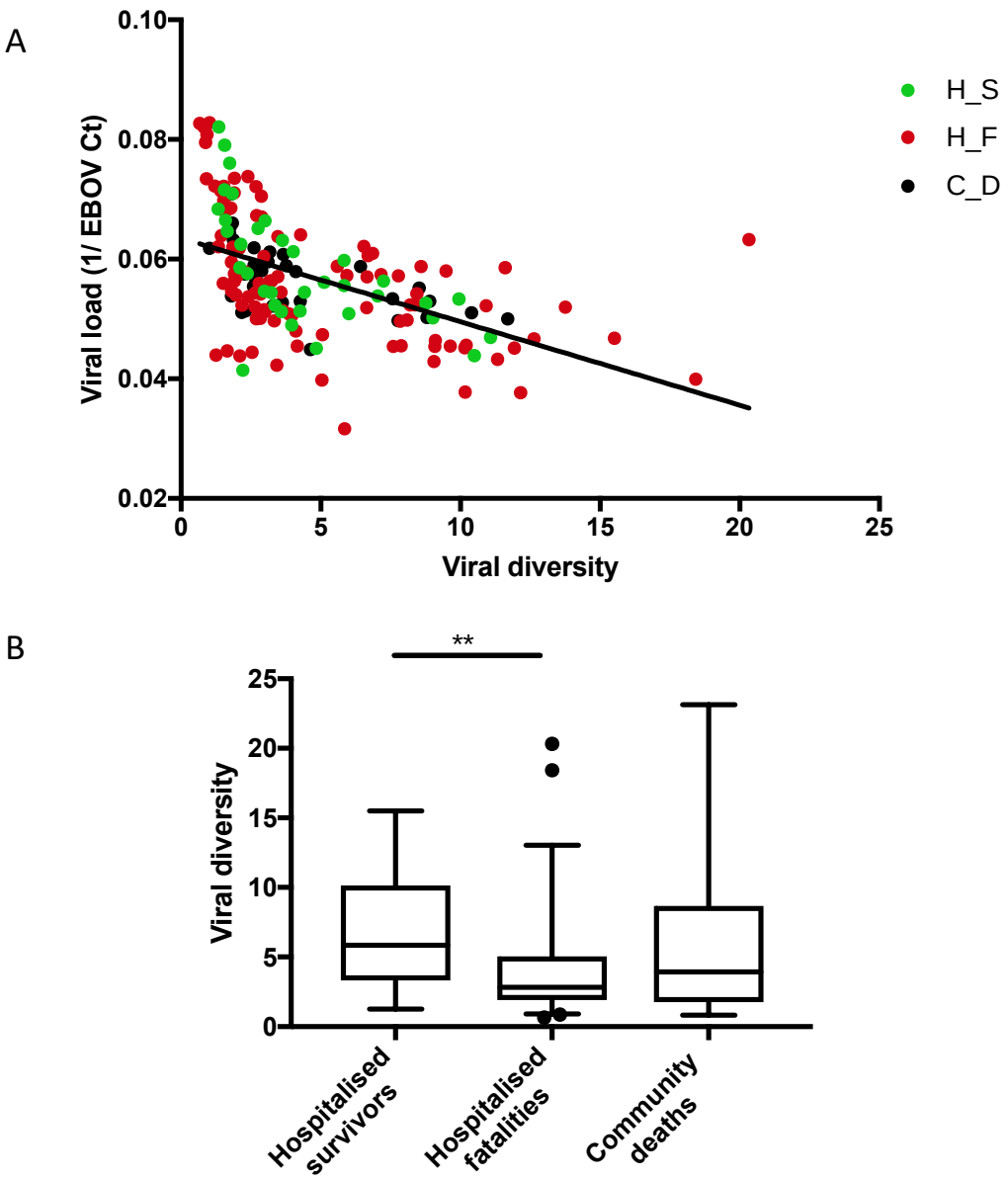


Figure 3.6: Relationship between viral load and viral genome diversity. *As viral load decreases, viral diversity increases (A). Points are colour-coded by patient group, where H_F are hospitalized fatalities (red), H_S are hospitalised survivors (green) and C_D are community deaths (black). Linear regression performed (black line),* $R^2 = 0.26$, $p < 0.0001$. *Viral diversity is significantly higher in acutely-infected survivors compared to fatalities (B). Viral diversity is a measure of the number of minor variants (at a rate of 0.05 or above) using a sliding window of 100 nucleotides along the length of the EBOV genome. One-way ANOVA performed with Holm-Sidak's corrections for multiple comparisons, ** p = 0.0019.*

3.2.4 Variation along the EBOV genome

After identifying some fundamental differences in viral populations in fatal and non-fatal EVD patients, the next step was to investigate where along the EBOV genome this variation was occurring. A number of RNA viruses have hyper-variable regions that are postulated to help in the response to selection pressures – particularly for immune evasion (Cane *et al.*, 1994; Delisle *et al.*, 2012; Guan *et al.*, 2012). Therefore, investigation into whether the intra-patient variation occurred in discrete regions along the EBOV genome was performed. Here, the proportion of minor variants at a rate of 0.05 or above were calculated as a proportion of the total coverage. This is shown as an average over a sliding window of 100 nucleotides (Figure 3.7). These data indicated that variation along the EBOV genome occurs in coding and non-coding regions alike. There appear to be clusters of variation at discrete nucleotide positions, which differ in hospitalised survivors (Figure 3.7A) in comparison to fatalities (Figure 3.7B). As suggested previously, the variation appears to be higher in survivors, with peaks of variation occurring in viral genes NP and L. Some of the variation along the EBOV genome is conserved between individuals, leading investigation into whether any mutations were associated with patient outcome.

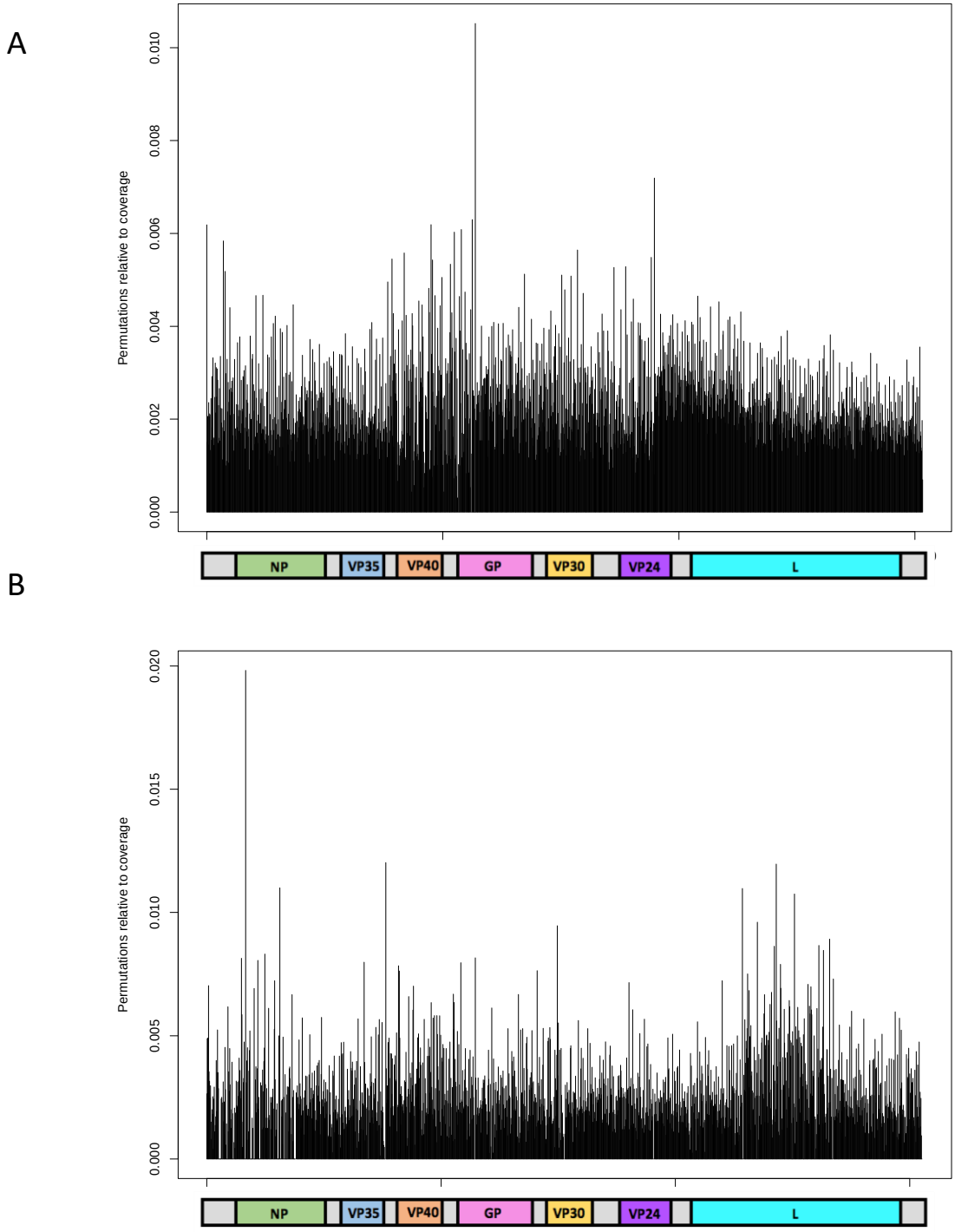


Figure 3.7: Intra-patient nucleotide variation across the EBOV genome. *Shown for hospitalised fatalities (A) and survivors (B). The ratio of minor variants to coverage in a sliding window of 100 nucleotides is plotted against nucleotide position.*

When the intra-patient nucleotide variation of survivors and fatalities was compared more specifically, several coding changes were identified to be maintained in survivors but not as readily in their fatal counterparts. The viral polymerase (L protein) was examined in further detail as this is seen to be an area of high variability in survivors of EVD (Figure 3.7B). Here, the most common mutations found in the samples were compared between survivors and fatalities of EVD (Figure 3.8; identified by Carroll *et al.* 2015). In most of the amino acid changes, the proportion was seen to be higher in hospitalised fatalities compared to survivors. Although the variation is higher in survivors, the specific mutations associated with amino acid changes (nonsynonymous changes) appear to be more common in fatalities.

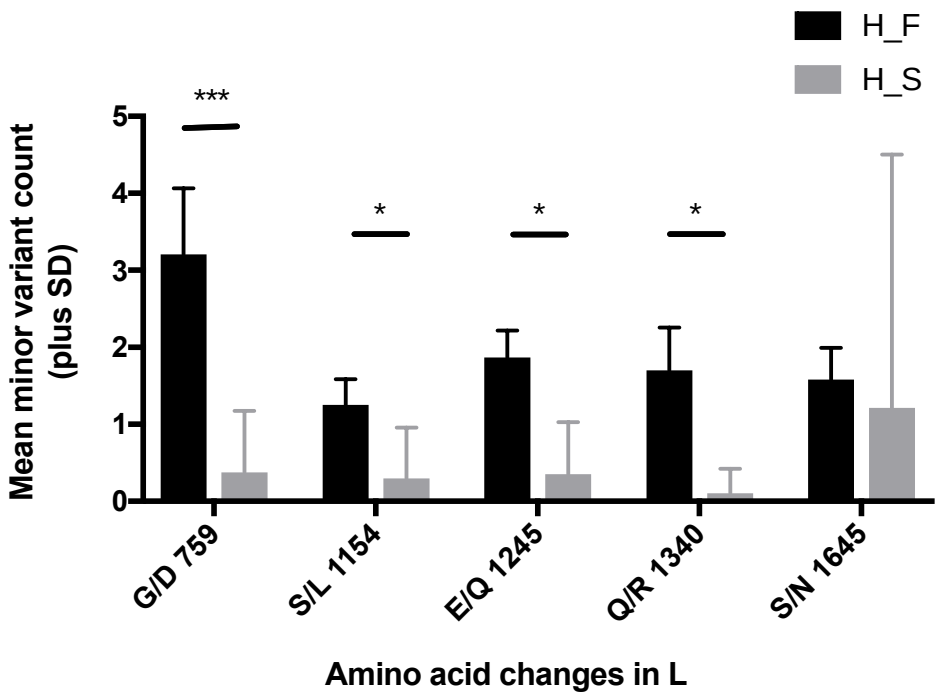


Figure 3.8: Certain amino acid changes in L are more frequent in fatal cases of EVD. *Multiple t tests were performed to compare hospitalised fatalities (H_F) and hospitalised survivors (H_S). * $p < 0.05$, *** $p < 0.0005$ with Bonferroni corrections for multiple comparisons. Here, counts of viral genomes with a specific amino acid change are normalised to the total number of samples in each group and presented as mean minor variant count plus standard deviation.*

To determine whether genetic variation of EBOV was confined to specific classes of nucleotide changes, all 12 nucleotide permutations (transitions and transversions) were determined for all patients, split by patient outcome (Figure 3.9). These data indicated that the A to G permutation was the most prevalent, with this effect being even more pronounced in hospitalised survivors. This permutation is suggestive of ADAR involvement. These enzyme binds readily to double-stranded (ds)RNA and are capable of RNA-editing via post-transcriptional modifications (Samuel, 2011).

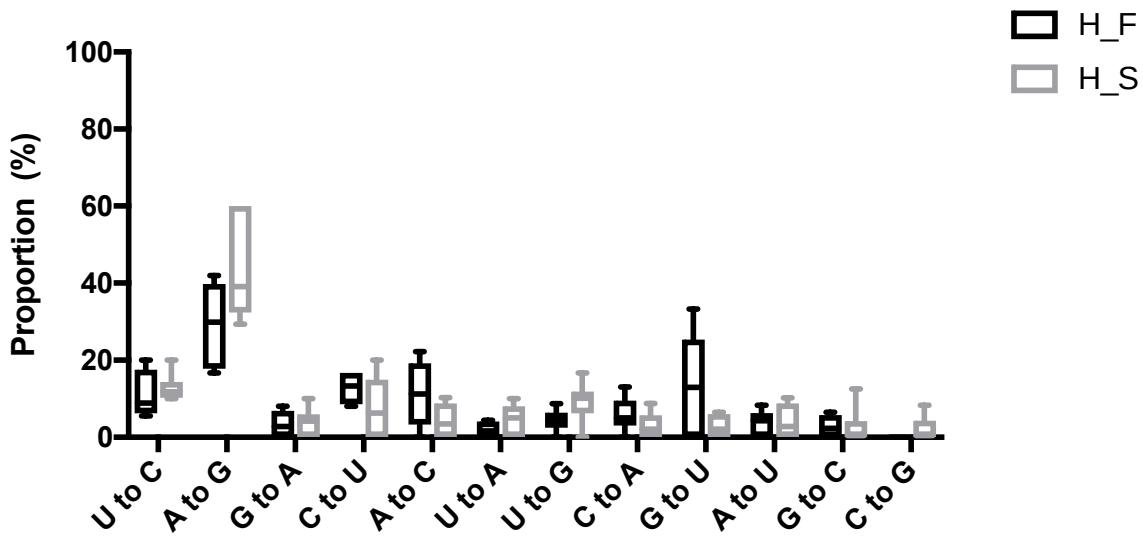


Figure 3.9: Intra-patient variation for all individual nucleotide permutations. *Intra-patient nucleotide variation for each of twelve possible permutations determined as a proportion of all minor variants. Where H_F are hospitalized fatalities (black) and H_S are hospitalised survivors (grey).*

The was no differential expression of ADAR between fatalities and survivors, which is perhaps unsurprising due to the potential intrinsic nature of this protein. However, this could be due to the fact that ADAR exists in two isoforms: ADAR1 (interferon-inducible) and ADAR2 (constitutively expressed in humans) (Samuel, 2011). Thus, transcriptomic data alone may not be able to accurately describe any difference in ADAR abundance.

3.2.5 Comparison of EBOV variation in different body fluids

One of the advantages of following up patients throughout their convalescence was the ability to sequence multiple samples from the same individual. Figure 3.10 shows a comparison of minor variants detected along the EBOV glycoprotein in a blood sample and semen sample from a convalescent patient. This gene was chosen due to the fact that it is the area of the highest risk of mutations. An interesting aspect of this comparison was that the same minor variants were not found in these different body compartments. However, this is not wholly surprising as these samples were taken seven days apart. The minor variants detected in the blood sample are more frequent but at a lower rate. However, once read depth is controlled for, the frequency of minor variants is not different between the two samples. Additionally, the ratio of synonymous to non-synonymous mutations is also not significantly different between these samples.

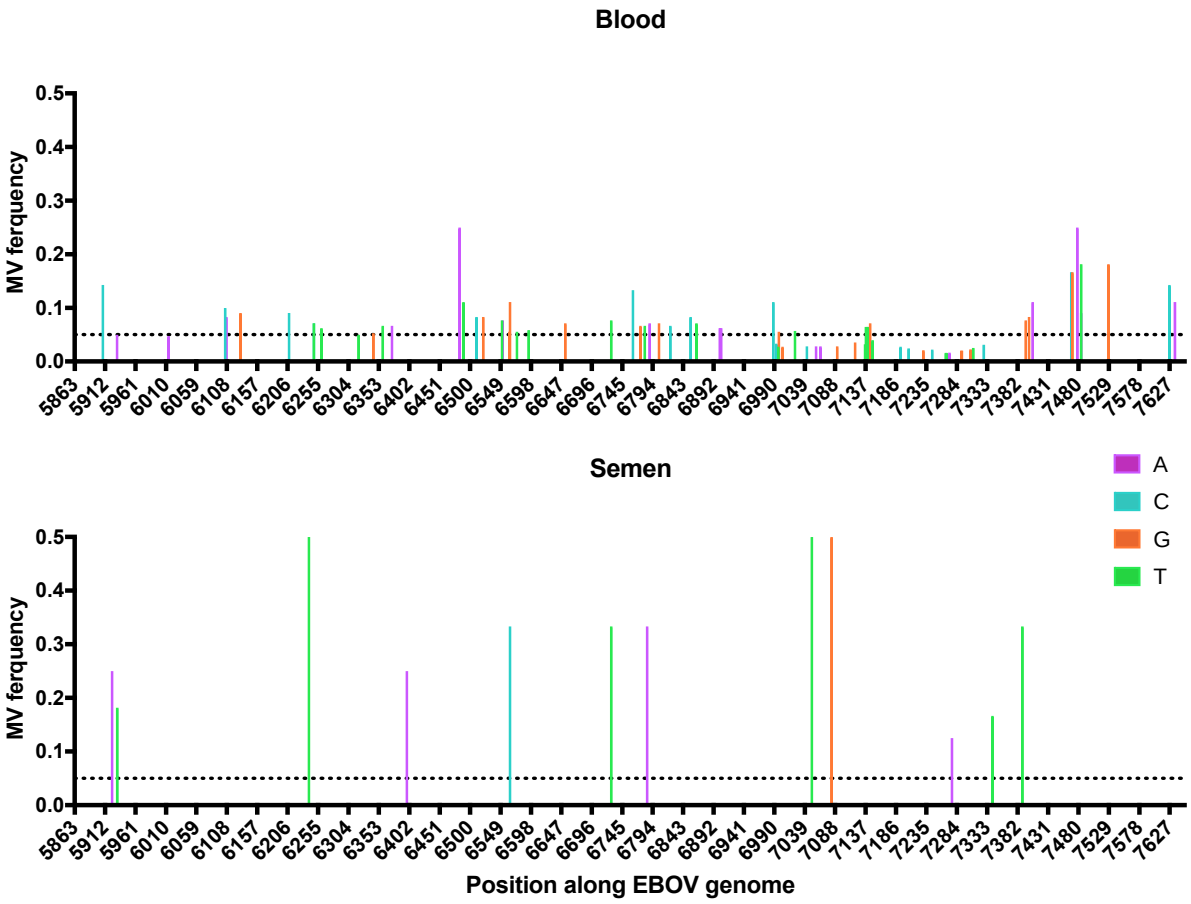


Figure 3.10: Glycoprotein variation of blood and semen samples from the same individual. *Instances of minor variants (MV) are shown across the EBOV glycoprotein gene, with the colour showing which base was coded for in that position.*

3.2.6 Identification of potential defective EBOV genomes

In order to optimise further analytical approaches into the investigation of EBOV, samples from infected THP-1 cells were sequenced and investigated. These macrophage-like human cells are a model cell line for the *in vitro* study of EBOV activity due to the evidence suggesting that these viruses are capable of infecting macrophages (Geisbert *et al*., 2003). Differentiated THP-1 cells were infected with EBOV-Makona (H.sapiens-wt/GIN/2014/Makona-Gueckedou-C05) – a near-clinical isolate – in CL4 conditions. Following infection, cellular and viral RNA was extracted 24, 48 and 72 h.p.i. These samples were sequenced using an Illumina HiSeq 2500. During mapping to EBOV using TopHat2 (Kim *et al.*, 2013), a fusion search was enabled in order to search for truncated genomes and potential DIs. Increasing numbers of sequence reads mapped to the EBOV genome across the time course, including the untranslated regions (Figure 311A). Reads mapped to all viral genes in varying abundance with fewest corresponding to the L gene, as expected from the sequential nature of transcription in the *Mononegavirales* (Figure 3.11B, 3.11C). Somewhat surprisingly, more reads mapped to the glycoprotein (GP) region than any other gene. This could indicate that this region or the mRNA was more abundant in the cell. Alternatively, it could suggest increased stability of GP, or identify an enrichment during the processing steps for sequencing.

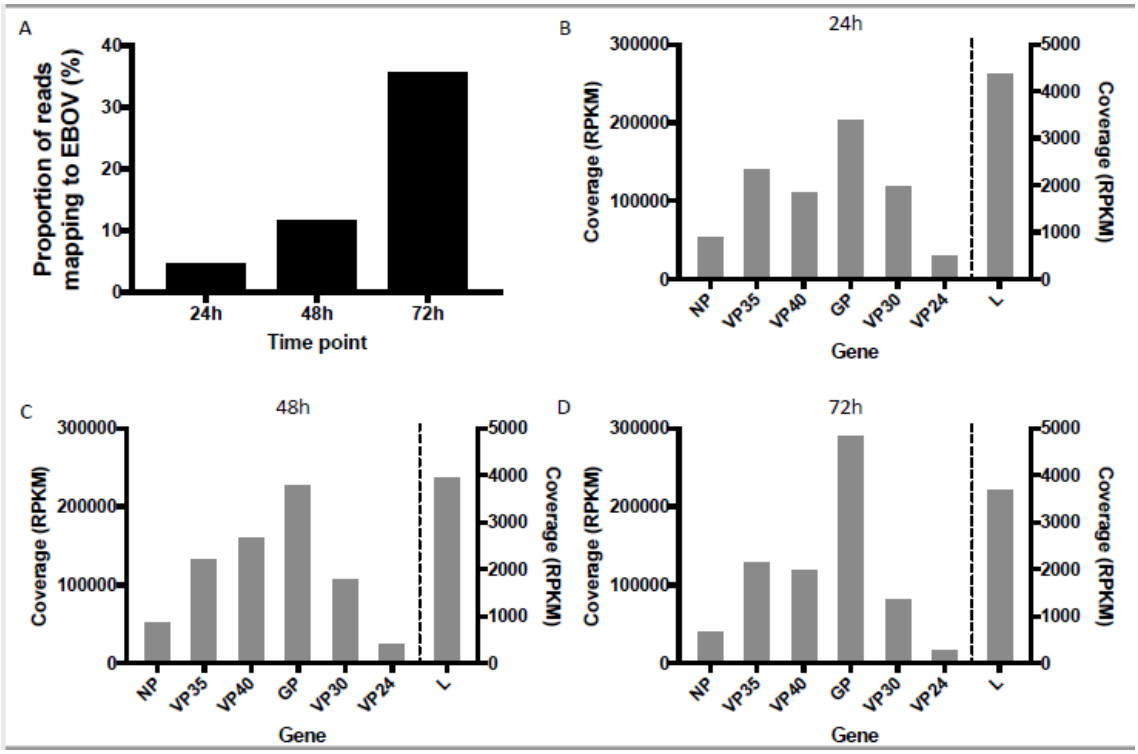


Figure 3.11: Infected THP-1 cells accumulate virus over time, however, the viral profile appears to remain constant. *After the RNA was extracted from infected cells and sequenced the proportion of reads mapping to EBOV increased over time, as calculated from the total reads (A). The coverage of viral genes stays relatively constant across the time course at 24 (B), 48 (C) and 72 (D) h.p.i..*

Searching for DIs in samples using RNA-seq was achieved by interrogating fusion events between disparate parts of the viral genome. Applying this approach to EBOV-infected THP-1 cells identified 193, 711 and 1625 fusion events at 24, 48 and 72 h.p.i respectively (Table 3.2, Figure 3.12). This value is proportional once the increased number of reads mapping to EBOV are considered. In Figure 3.12 the section of the EBOV genome that has been deleted from the potential DI is shown as a horizontal line, with the depth of the colour indicating the level of support for this fusion event. These apparent discontinuities could have arisen during the reverse transcription step for sequencing or directly from the EBOV L protein. For this reason, the number of sequence reads supporting each fusion event was examined. Fisher's exact statistic was used to create a stochastic threshold of 6 supportive reads, reasoning that rare events were less likely to be biologically significant. This greatly reduced the number of possibly relevant DIs from 2398 to 225. Of these, the ones most likely to be functioning as DIs are those appearing in each time point and increasing in abundance. This was down selected to nine serious candidates, with locations tending towards the 3' end of the EBOV genome.

Table 3.2: Detection of fusion events in EVOV-infected THP-1 cells. *EBOV read depth is the number of reads mapping to the EBOV genome.*

	24h	**48h**	**72h**
EBOV read depth	958334	2273909	7649409
Fusion events detected	193	711	1625
Rate of fusion events (-Log[% fusion events])	3.70	3.50	3.67

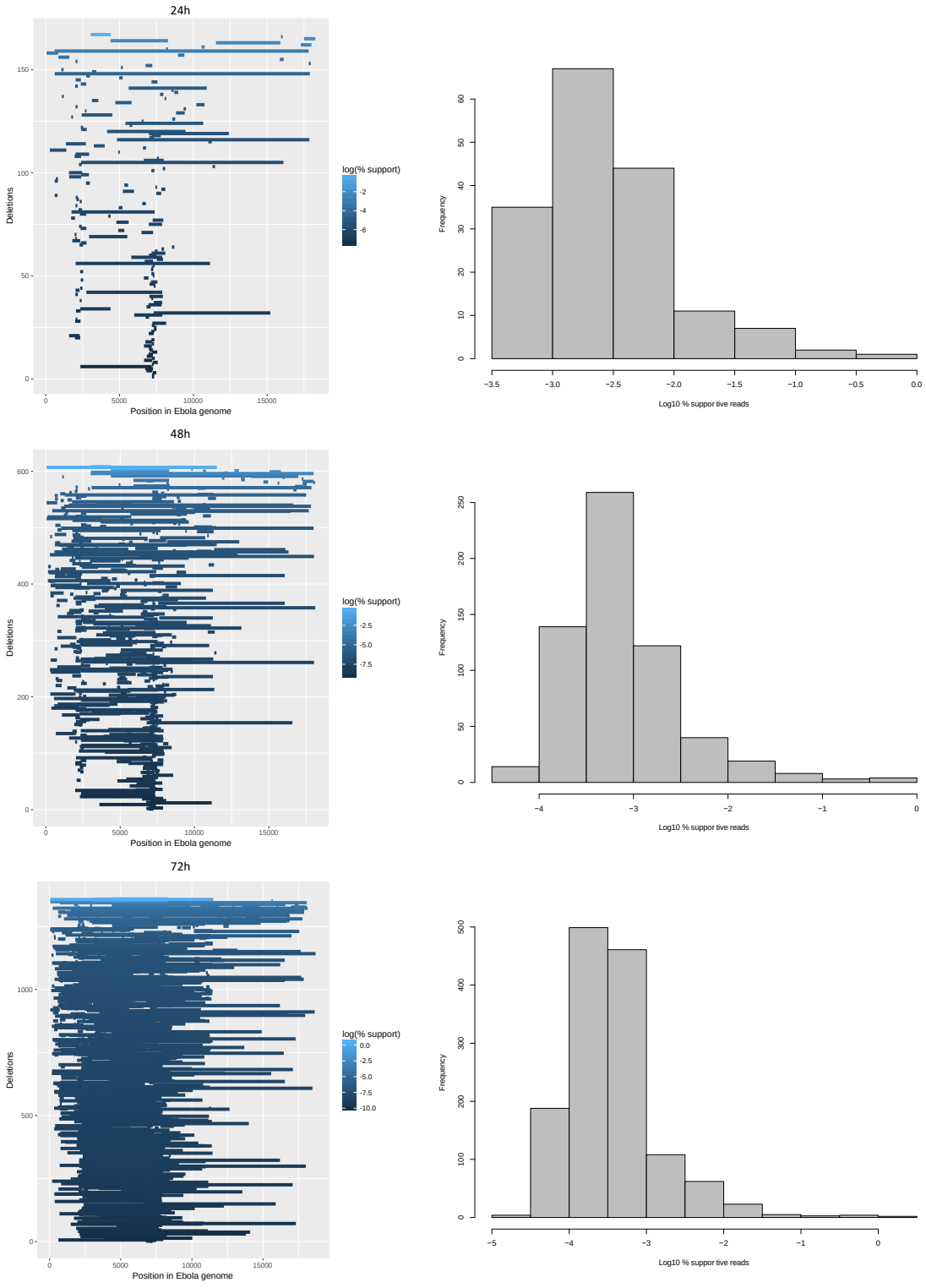


Figure 3.12: Identification of potential defective genomes in Ebola-infected THP-1 cells. *Infected cells were harvested at 24 (A), 48 (B) and 72 (C) hours post infection and fusion events detected following Illumina RNA-sequencing. The colour of each deletion event depicts how frequently it was supported in the RNA-sequencing analysis, with high levels of support being light blue and lower levels of support being dark blue. The histograms on the right-hand side show the levels of support for all the fusion events at the corresponding time point.*

After successfully identifying these fusion events in RNA extracted from experimentally-infected cell lines, the next step was use this pipeline to search for the same in clinical samples from the West African outbreak. Upon doing so, very few fusion events were found, even when a range of sample types were investigated (Table 3.3, 3.4). Of those detected, very few were of substantial length, and likely to transmit any kind of advantage in speed of replication over the authentic virus. A number of fusion events were detected in blood and swab samples with no fusion events found in the urine or semen samples from convalescent patients. These samples generally had a lower viral load than the blood or swab samples we procured, therefore potentially not providing a sufficient coverage for fusion events to be detected. In the blood and swab samples, the vast majority of fusion events were suggestive of defective genomes that were shorter than the full EBOV genome by 100-300 nucleotides. Deletion events this small are unlikely to confer any kind of evolutionary benefit and therefore are of little interest. Some longer deletions were detected with the potential of being DIs. However, of these, only a smaller number would be considered to be "true" events due to the low rate of detection of these fusion events.

Table 3.3: Summary table of fusion events found in patient samples. *Only blood samples with very high Ct values were considered to be appropriate for this analysis, hence the reduced number of samples involved.*

Sample type	Number of samples	Fusion events detected	Fusion events with at least 6 supportive reads
Swab	28	82	7
Blood	44	18	0
Semen	16	0	0

Table 3.4: Fusion events found in clinical samples. *All of these samples are oral swabs from deaths in the community. These represent potential deletion DIs with length of deletion and confidence in fusion event shown.*

Patient ID	Start of fusion	End of fusion	Deletion length	Reads that support the fusion	Reads that don't support the fusion	Proportion of support (%)	Patient EBOV Ct
28	18035	18189	154	6	416	1.42180095	19.47
74335	18025	18179	154	7	2944	0.23720773	14.1
958	319	2510	2191	6	1323	0.45146727	16.18
958	6293	6424	131	34	42951	0.07909736	
958	6305	6434	129	11	48642	0.02260909	
958	18025	18179	154	9	13568	0.06628858	
958	18042	18191	149	11	8370	0.13124925	

3.3 Discussion

The identification of certain patient groups most at risk of a fatal outcome of EVD is critical in patient management. Here, and in other studies (Fitzpatrick *et al.*, 2015; Vega *et al.*, 2015; Kerber *et al.*, 2016) and previous outbreaks (Towner *et al.*, 2004), viral load is identified as being an indicator of patient outcome, as well as those with malaria as a co-infection (Figures 3.1, 3.3). In this study, age did not appear to impact prognosis, which opposes previous investigations (Fitzpatrick *et al.*, 2015). The explanation for this could simply be that the vast majority of patient samples were taken from individuals aged between 20 and 60, thus older and younger individuals are not well represented. A higher proportion of women appeared to die in comparison to men, though this was not statistically significant. During the most recent EVD outbreak in the Democratic Republic of Congo, more women have seen to be infected than men (Nkengasong and Onyebujoh, 2018). This can potentially be attributed to the role of women in West Africa. Acting as caregivers, these individuals are at higher risk of infection and potentially wait longer before seeking the care or treatment they require, whilst still attending to other members of their family.

This study demonstrates the importance of nucleotide diversity for EVD management. Next-generation sequencing allows the monitoring of EBOV evolution in the human host and demonstrates the magnitude of intra-patient variation. It is assumed that this intra-patient variation is, at least in part, driven by host defence mechanisms that can cause hyper-mutation of the RNA genome. Potential manipulation of the mutation frequency of EBOV during replication may provide a therapeutic avenue that could tip the balance between a fatal and non-fatal outcome.

Here, a higher viral diversity was observed in survivors of infection compared to fatalities. This could be indicative of a more effective immune response mounted by the survivors. A robust host immune response can lead to hypermutation in viruses as a method of immune evasion. This potentially leads to error catastrophe and a decline in the replication rate or pathogenicity of the virus.

A to G transitions have been observed in a number of viral infections, including EBOV and other filoviruses (Zahn *et al.*, 2007; Khrustalev *et al.*, 2015; Whitmer *et al.*, 2018). A pattern of substitution has been identified in an analysis of patient samples taken from seven months of the outbreak in Sierra Leone (Park *et al.*, 2015). Thus, the increased frequency of A to G is common to different geographical isolates of the virus. It is notable that during the course of mouse adaptation, Marburg virus (MARV) underwent apparent ADAR editing, with clusters of A to G changes accumulating in the genomic RNA (Lofts *et al.*, 2011; Qiu *et al.*, 2014). Potential A to G transitions have also been observed in EBOV (Mayinga) and MARV RNA in cell culture experiments (Shabman *et al.*, 2014). The A to G transition has also been observed, and therefore likely to contribute to the generation of escape mutants, from infected NHPs treated with a cocktail of three monoclonal antibodies targeted to the EBOV glycoprotein (Qui *et al*., 2016). This A to G transition is the most common among the patient samples when compared to a Makona isolate from very early in the outbreak (Figure 3.9). This is potentially due to the activity of ADARs – notably ADAR1, an interferon (IFN)-inducible protein. The activity of ADAR1 has been observed to have potential anti- and pro-viral activity (Samuel, 2011).

The following up of patients throughout their convalescence has allowed the study of individuals experiencing persistent infections. The mechanism behind EVD persistence is not well understood (Varkey *et al.*, 2015; Sissoko *et al.*, 2017). The presence of DIs is one potential explanation, and these were readily identified in experimentally-infected THP-1 cells. These were seen to accumulate over time (Figure 3.12). Of those identified, the majority were only partially truncated and therefore unlikely to confer much of an evolutionary advantage over a wildtype virus. Studies on influenza have shown that the introduction of synthetically constructed DIs can dampen viral replication, thus demonstrating their potential use in a therapeutic setting (Dimmock *et al.*, 2008). However, in this dataset very few were identified in patient samples. This is particularly troubling as samples from individuals with persistent infections were analysed. The explanation for this could be due to the lack of sufficient coverage in many patient samples, which is particularly notable when compared to the *in vitro* study, where high multiplicity of infection (MOI) can be used to infect cells. These high MOIs often do not reflect accurately a natural viral infection. Equally, this could be suggestive that these truncated genomes are simply an artefact of the replication of RNA viruses, and unlikely to be an important factor in EVD persistence in humans. A supporting factor of this is the fact that very little damage occurs in these persistently infected body compartments, suggesting that the virus has somehow been temporarily "turned off".

Here, Nanopore and Illumina RNA-seq could be closely compared. Unfortunately, the partially degraded condition of the RNA extracts proved to be a substantial issue when employing a MinION using Nanopore sequencing. There was a definite bias towards the 5' end of the genome, which was expected, however this left the 3' end almost

without coverage (Figure 3.4A). These reads are much longer when compared to the 150 bp covered during Illumina sequencing. However, in this case, these long reads provide little advantage, as the overall coverage is much lower than samples sequenced using Illumina platforms. These shorter reads better allow issues of RNA quality to be overcome. No optimisation of this technique was performed and this MinION run was performed for cDNA sequencing following a PCR step, thus introducing a bias towards certain transcripts. Further bias was introduced by the fact that many of these reads begin at the same point along the EBOV genome. This could be due to the degradation of the sample that was sequenced. This bias is problematic when considering one of the aims of this project: the generation of a consensus genome. Direct RNA sequencing by Nanopore may return more informative results, and the error rate has dropped substantially over the last few years (Jain *et al.*, 2016; Garalde *et al.*, 2018). Additionally, the samples involved in this study were extracted in sub-optimal conditions at the ETC and stored for extended periods of time. Immediate sequencing following RNA extraction, and the use of an amplicon-based approach would maximise the data gained from Nanopore sequencing, especially when accommodating clinical samples, which could be of lower quality. The potential uses for this technology are plentiful and represent an exciting avenue for personalised medicine.

3.3.1 Future work

After the identification of a small number of potential DIs, the next step would be to assess their function *in vitro* in a number of EBOV-infected cell lines. This could give further indication as to whether these truncated genomes play an important role in

EBOV infection. Potentially, sequencing positive samples from convalescent patients at a higher depth could also allow such events to be uncovered.

Chapter Four: Delineating the host response in Ebola virus disease and the association with outcome

4.1 Introduction

4.1.1 Immune response to acute EVD

Analysis of acute sufferers of EVD, including four patients who received intensive treatment at Emory University Hospital (GA, USA) during the 2013-2016 outbreak in West Africa, indicated that EBOV infection elicited a robust immune response (Hunt *et al.*, 2015; McElroy *et al.*, 2015). This challenged previous *in vitro* studies that suggested EBOV was able to effectively suppress the immune system through the activity of T lymphocytes, DCs and macrophages (Bosio *et al.*, 2003; Mahanty *et al.*, 2003; Basler and Amarasinghe, 2009). During the 2013-2016 outbreak, a feature of the host response to EVD was found to be an overwhelming of the immune system. This was characterised by uncontrolled inflammation and lymphopenia (Ruibal *et al.*, 2016).

4.1.2 Predicting patient outcome of acute EVD

The ability to predict patient outcome of infection based on changes in abundance of host factors could provide guidance for possible treatments and potentially uncover new therapeutic strategies for EVD. This is particularly important in outbreak situations, where clustering of patients requiring similar levels of care could be

critically important to effective delivery of the best possible care by healthcare workers. Several promising biomarkers were identified in a previous outbreak of SUDV (McElroy *et al.*, 2014b, 2014a). However, currently, semi-quantitative viral load measurements by qRT-PCR (reported as a Ct value) are considered the gold standard for predicting EVD outcome. Viral load measurements for EBOV do correlate with outcome and therefore have some predictive ability (Carroll *et al.*, 2015; Fitzpatrick *et al.*, 2015). This has been implemented in the triage of patients receiving treatments, for example in Favipiravir trials during the 2013-2016 outbreak (Sissoko *et al.*, 2016). However, using viral load as a predictor functions less well where Ct values are intermediate, i.e. between Ct 20 and 22. Here the outcome was approximately equal between survival and a fatal infection.

4.1.3 Objectives and hypotheses

In order to characterise the infection of EBOV-Makona in the 2013-2016 outbreak, deep sequencing was used to define the transcriptomic profile of blood taken from acute patients who either went on to survive or die from EVD. These samples were obtained from patients located in Guinea as described previously. During this outbreak, Guinea experienced the highest mortality rate for EBOV infection (Chapter 1, Figure 1.3). Thus, EVD patients in Guinea were optimal for correlating changes in the host response to outcome as there was little influence from intensive palliative and/or experimental care. This supportive care was utilised heavily to treat repatriated healthcare workers in high-income countries such as the USA and UK (Uyeki *et al.*, 2016). In these environments, 81.5% of patients who received supportive care survived while in contrast, the survival rate in Guinea was approximately 40%. However, this figure does not take into account the many asymptomatic cases that were identified

during this outbreak (Akerlund *et al.*, 2015; Glynn *et al.*, 2017). This fatality rate varied with patient age, with a higher proportion of older and younger individuals dying in this outbreak. This could, therefore, potentially be linked to the enrichment of specific pathways by individuals of a similar immunological maturity.

The transcriptional signatures from acutely ill patients were compared to profiles obtained from convalescent patients who had recovered from EVD and were EBOV-negative by qRT-PCR. Data was also obtained from healthy volunteers. This was performed in order to highlight differences in the host response to EVD at the time of acute infection with respect to patient outcome. These investigations could allow the development of prognostic tools, allowing patient outcome to be predicted accurately at the point of diagnosis at the ETC.

4.3 Results

In order to identify changes in the host transcriptome of individuals infected with EBOV, the samples described in Chapter 5, Table 5.1 were utilised and down selected. Here, samples were selected that had a similar within-sample transcriptional profile (with an average correlation co-efficient above 0.8), thus reducing extreme variation within one patient group due to mRNA quality issues. Due to the diagnostic nature of the sample taking, these were from a mixture of *Plasmodium spp.* (the causative agents of malaria) negative and positive patients. Quality control criteria were used to identify and remove datasets from downstream analysis that showed evidence of having degraded mRNAs (expected from field sample collection). These criteria included the removal of samples that showed poor mapping to the human genome and low correlation values. From an initial set of 170 sequenced samples from individual patients, application of these selection criteria led us to discard 29 and analyse the data from 141 unique patients: hospitalised_survivors (n = 52) and hospitalised_fatalities (n = 89) (Table 4.1). The clinical outcome of these patients was unknown at the time of sample collection and subsequently recorded as they either succumbed to a fatal infection or survived EVD.

Table 4.1: Summary of patients involved in this study. *These have been split according to sample type and patient outcome. Here, patient age and EBOV Ct are described by the range and mean. N describes the total number of samples in each patient group.*

Patient group		**Hospitalised fatalities**	**Hospitalised survivors**	**Convalescent controls**	**Community deaths**
N		89	52	16	24
% Male		48.3	29.0	93.8	29.1
Age (years)	Range	2 mo – 77	10 mo – 68	18 – 40	1 - 70
	Mean	24.7	33.7	33.1	30.0
EBOV Ct	Range	12.1 – 26.6	15.8 – 31.6	NA	12.4 – 31.9
	Mean	17.0	21.9	NA	18.8

4.3.1 Analysis of gene responses that differentiate patient outcome during EVD

The transcriptomes from the blood of acutely ill EVD patients were analysed to identify mRNAs in both survivors and fatal cases that showed greater than 2-fold changes (at a false discovery rate [FDR] of 5%) in abundance compared to a convalescent control group. This control group provides a baseline for gene expression. Blood samples were taken from a group of 16 EBOV-negative individuals (by qRT-PCR) and convalescent for the disease (Table 3.1). During the 2013-2016 outbreak, amid the breakdown of the in-country health-care system and the stigma of being associated with ETCs, acquiring samples from non-infected patients was challenging. These survivors represented a known EBOV-negative population that were also tested and confirmed to be *Plasmodium*-negative with no other overt clinical signs of an acute infection. A historical dataset was also used, which was obtained from the RNA-seq analysis of peripheral blood taken from healthy volunteers based in British Columbia (Canada) (n = 6) and thus would not have been exposed to EBOV or a range of other pathogens present in West Africa (Shin *et al.*, 2014) (GEO Number GSE53655). Here, the transcriptomes of the convalescent group were compared to the healthy volunteers and not shown to be significantly different (Liu *et al.*, 2017). Therefore, the convalescent control group was used as a comparison due to the likely similarities in host genomes.

In the hospitalised survivor (H_S) group, transcripts from 437 genes were increased in abundance compared to the convalescent controls. This was increased to 1658 in the hospitalised fatalities (H_F), with 370 gene transcripts being shared between the two groups (Figure 4.1). Of these, many were transcripts of pro-inflammatory factors including CXCL10, CCL2, CCL8 and CXCL11, which were found to be of higher

abundance in fatalities compared to survivors. This phenomenon has also been observed in non-human primates (Rubins *et al.*, 2007; Caballero *et al.*, 2016).

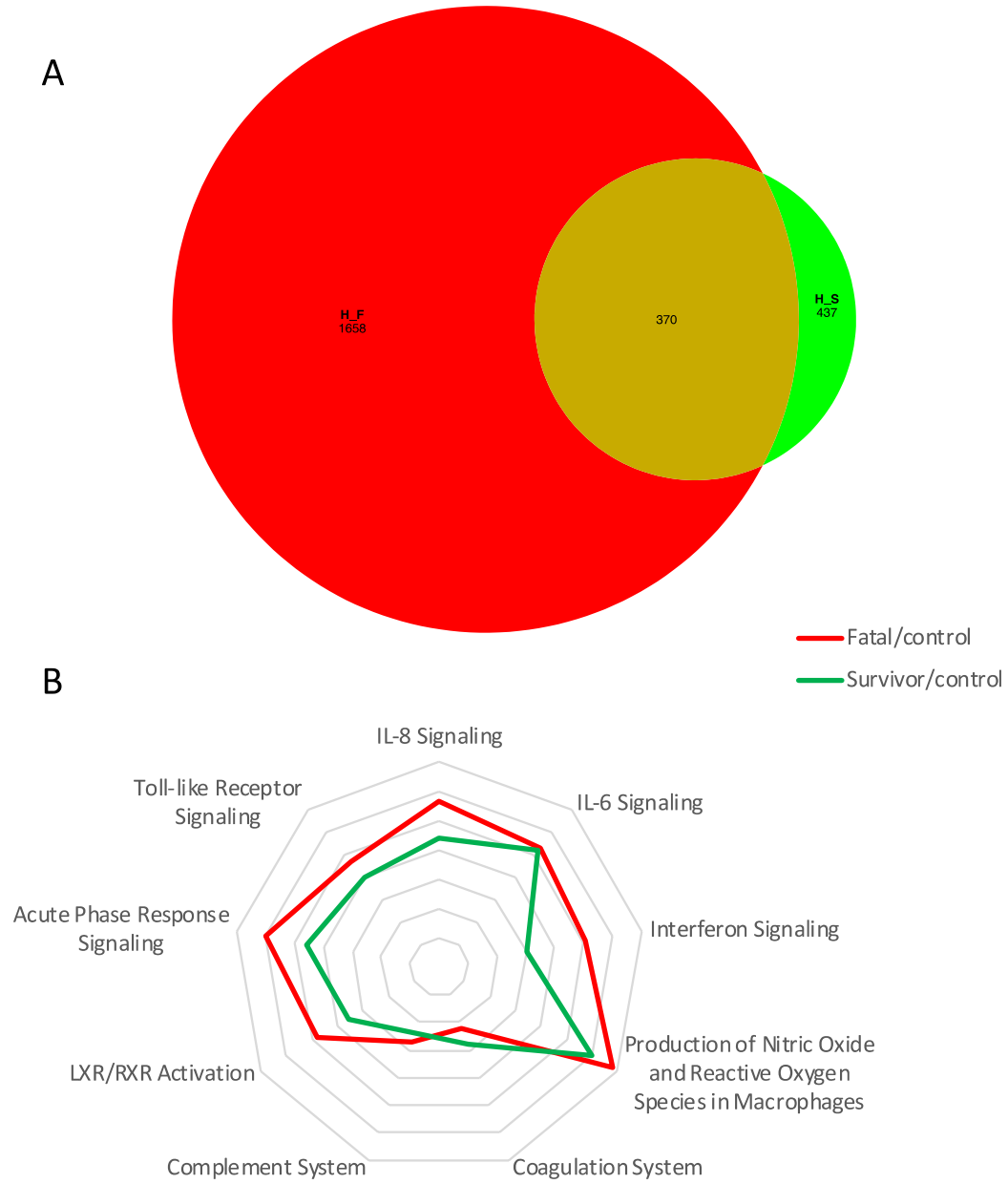


Figure 4.1: Transcriptional responses to acute EVD during the 2013-2016 West African outbreak. *Venn diagram depicting genes that are differentially expressed from fatalities to controls (red) and survivors to controls (green) (A). Radar diagram depicting pathways that are all upregulated in acute EVD sufferers verses controls, split into fatalities and survivors, showing the z-score generated by IPA (B).*

There were distinctive differences in gene expression patterns when the transcriptomes from hospitalised fatalities were compared to those of hospitalised survivors. This was most strongly observed in genes associated with coagulation and acute phase signalling; with the top ten differentially expressed (DE) genes between hospitalised individuals and convalescent patients shown in Figure 4.2. In these cases, the gene expression levels are higher in fatalities compared to survivors as well. Four of the DE genes associated with the clotting cascade were fibrinogen alpha chain (FGA, logFC +14), beta chain (FGB, logFC +13) and gamma chain (FGG, logFC +11). Later stages of EVD are often associated with coagulation disorders, with severe cases more likely to experience coagulopathy (Fletcher *et al.*, 2014). Although overt haemorrhage was not often reported in this outbreak, those with abnormalities of the coagulation system were reported to experience more severe clinical symptoms (McElroy *et al.*, 2016; Lanini *et al.*, 2018).

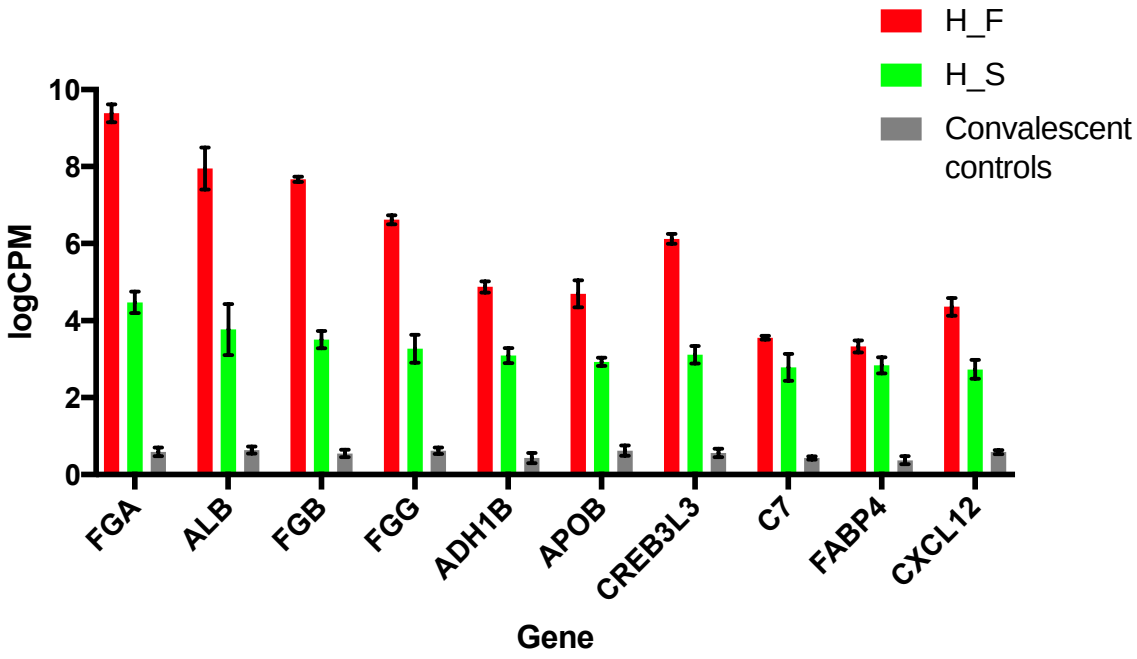


Figure 4.2: Top differentially expressed genes comparing acute sufferers of EVD, separated by patient outcome. *Here, log values of the normalised counts per million are shown, adjusted for feature length. Hospitalised survivors (H_S) are shown in green (n = 52), fatalities (H_F) in red (n = 89) and convalescent controls in grey (n = 16).*

Interestingly, many of these top DE genes are involved in the same pathways of organ injury and abnormal function, as identified using IPA. Of these, the majority are increased in abundance in fatalities compared to survivors, and is enriched in hospitalised fatalities. This correlates well with reports of individuals suffering end-stage EVD experiencing major organ dysfunction (Fletcher *et al.*, 2014). In order to emphasise this, genes associated specifically with renal damage are highlighted with a pink border (Figure 4.3).

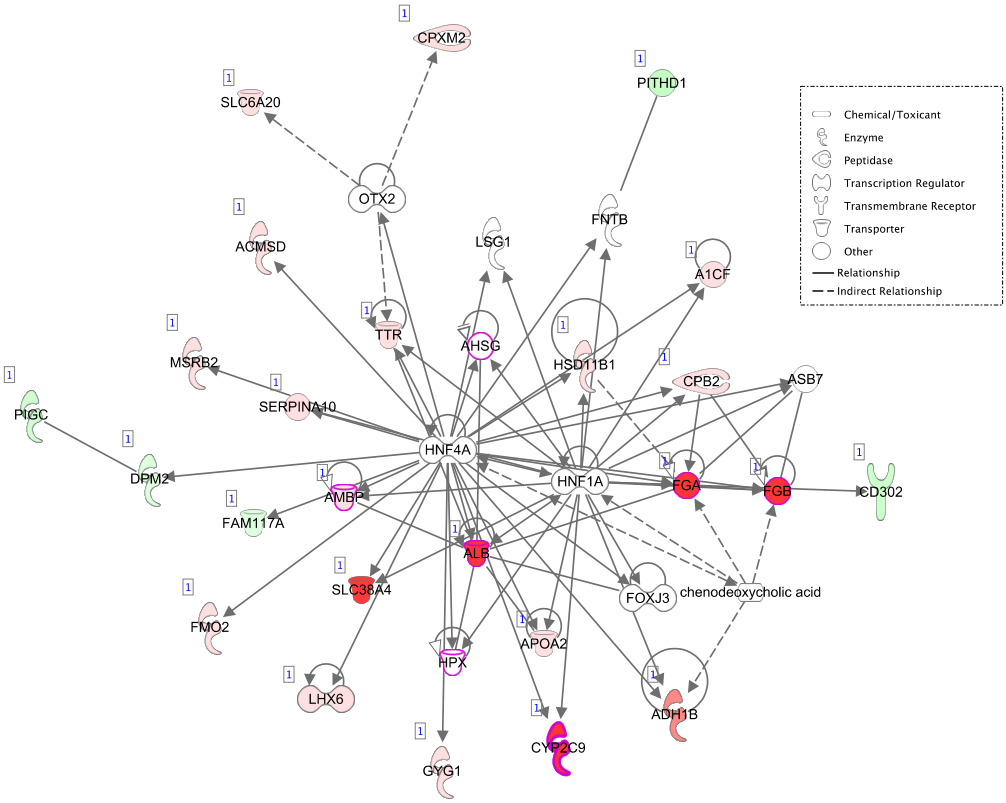


Figure 4.3: Network of genes involved in organismal injury and abnormalities. *Those genes that are upregulated in H_F are depicted in red, while downregulated genes appear in green. Genes outlined in pink are specifically associated with renal damage. Relationships quantified using Ingenuity Pathway Analysis (IPA, Qiagen).*

In order to validate these findings, primers were designed for three of these top genes and qRT-PCR performed using a CFX96 (Bio-Rad) following amplification of cDNA using random hexamers. RNA extracts from the blood of six hospitalised survivors and six fatalities, not used in the previous DGE analysis, were tested for the presence of FGA, C7 and APOB. Here, the Ct values were normalised to RPL13A, a housekeeping gene that was seen to not be significantly different across patient groups following RNA-seq. As in the RNA-seq analysis, mRNA abundance of these genes was higher in hospitalised fatalities when compared to survivors, showing an increase in the abundance of FGA, C7 and APOB in fatal cases of EVD (Figure 4.4A). Additionally, a PCR array was performed covering "Human Inflammatory Cytokines and Receptors" (Qiagen) (Figure 4.4B). This profiler array assesses the expression levels of 84 genes associated with the inflammatory immune response (see Appendix 7.3 for full list). A further 12 independent samples were used to again compare the two patient groups. The PCR array includes a number of housekeeping genes in order to effectively normalise samples. These normalised values were then compared and the 84 potential genes of interest plotted against each other. Out of those genes, two were significantly upregulated in H_F compared to H_S: IL-8 and CXCL12. This fit well with the findings of the Illumina analysis. The expression levels of the vast majority of genes were not significantly differentially expressed between the two groups, however, this is likely due to the small sample size. In this analysis, CD40L was observed to be downregulated in H_F compared to H_S.

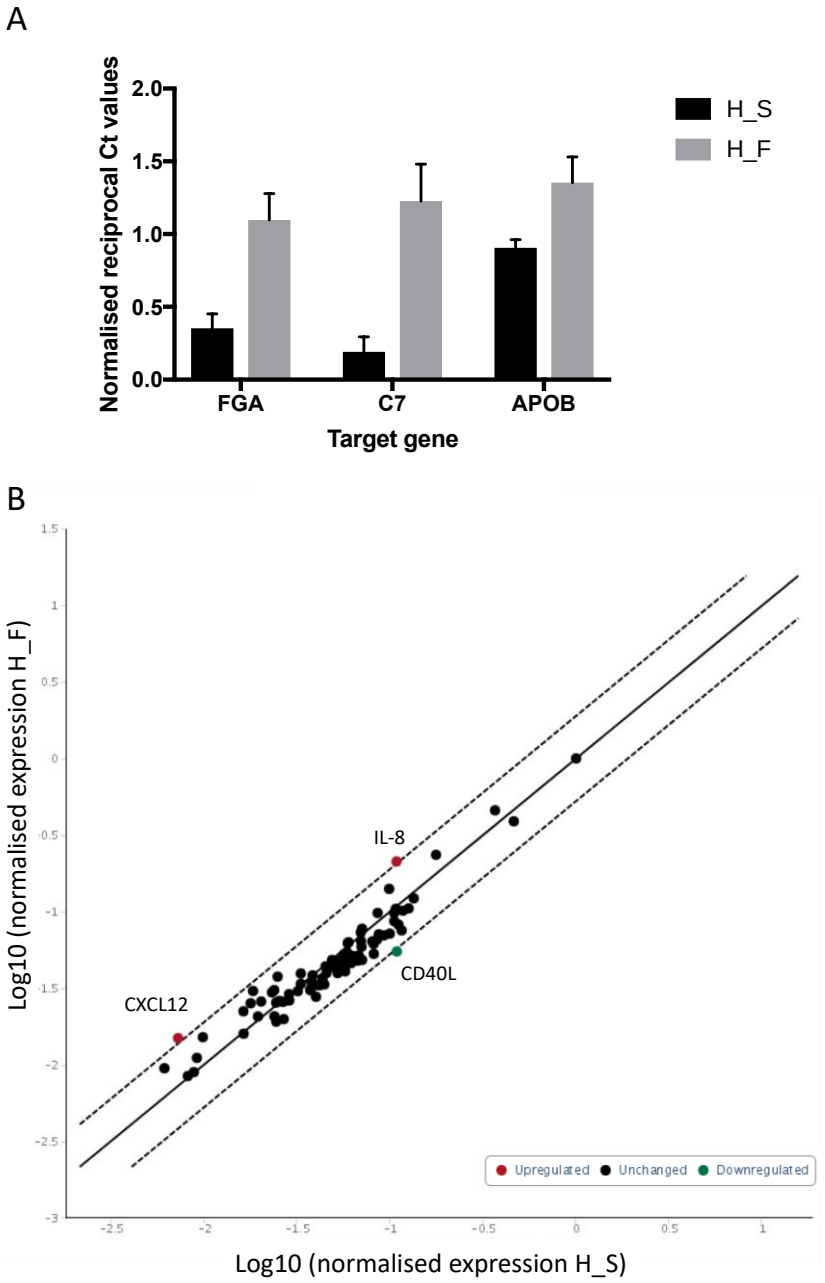


Figure 4.4: Comparison of hospitalised surivvors and fatalities using qRT-PCR techniques. S*ix hospitalised fatalities (H_F) and six hospitalised survivors (H_S) support the findings of the RNAseq analysis using targeted qRT-PCR of FGA, C7 and APOB (A). These were independent of the samples involved in the RNA-seq analysis. Values were normalised to housekeeping gene RPL13A. A profiler PCR array highlights the differential abundance of CXCL12, IL-8 and CD40L in these samples (B). Of the 84 genes associaed with the inflammatory response CXCL12 and IL-8 are shown to be upregulated in these RNA extracts from fatal infections.*

4.3.2 Analysis of gene pathways differentiating patient outcome during EVD

Pathway analysis using the KEGG database showed that hospitalisation of EVD patients was associated with a significant enrichment of genes from within the same signalling pathways, as compared with non-hospitalised patients. The most significantly represented included gene sets associated with interferon signalling, complement pathway, coagulation and cytokine-cytokine receptor interaction. Many interferon-stimulated genes (ISGs) were strongly increased in abundance in all acute infection cases compared to the convalescent control group. This potential link between IFN expression and immunopathology has been noted in the 1995 Kikwit outbreak of EVD in the Democratic Republic of Congo (Villinger *et al.*, 1999). In order to further investigate, these data were also analysed using Ingenuity Pathway Analysis (IPA, Qiagen Bioinformatics) to confirm the identity of signalling pathways associated with infection. This gave confidence to the KEGG analysis as many of the same immune pathways were identified, such as complement, IFN signalling and the coagulation cascade (Figure 4.1B). Further pathways were also identified that were indicative of a robust cytokine response – i.e. IL-6 and IL-8 signalling. In general, these pro-inflammatory responses appear to be enriched in hospitalised fatalities when compared to survivors (Figure 4.1B). High levels of both IL-6 and IL-8 have been flagged previously as being associated with fatal cases of EVD (Hutchinson and Rollin, 2007; Wauquier *et al.*, 2010).

4.3.3 Immune cell profiles of EVD sufferers

These observed changes in mRNA abundance in the blood of patients with varied prognoses could be due to the alteration of immune cell profiles. The altered

regulation and/or recruitment of specific immune cells could impact the transcriptomic signature of blood. Due to the constraints of this dataset, digital cell quantification (DCQ) was employed to identify which cell types may have been differentially abundant in hospitalised fatalities verses hospitalised survivors relative to the convalescent control group. This technique allows the immune cell profile to be predicted using only transcriptomic data via the implementation of a digital cell algorithm. The gene expression levels are utilised to forecast the potential abundance of cells types in each sample using FACS separating markers, which can then be compared across samples of a similar nature (Altboum *et al.*, 2014). In both acutely ill groups the predicted immune cell profile involves a decrease in CD4+ T cells, which has been previously observed in NHPs (Reed *et al.*, 2004). Conversely, CD8+ T cells are potentially increased in hospitalised survivors. The T cell response has been noted to be dysregulated in EVD (Ruibal *et al.*, 2016). This DCQ also demonstrated a predicted fall in the circulating monocyte population, with this being most significant in fatal infections (Figure 4.5). This has been supported by flow cytometry on further samples from the 2013-2016 outbreak. This analysis indicated that CD14+ classic monocytes were in lowest abundance in individuals who suffered a fatal infection when compared to EVD survivors and those with acute malaria who tested negative for EBOV (Liu *et al.*, 2017). NK cells were also predicted to be more abundant in survivors of EVD (Figure 4.5). This is consistent with the increase in mRNA abundance in hospitalised survivors of the NK markers IFNγ (4-fold higher than convalescent controls) and perforin (3-fold higher than controls).

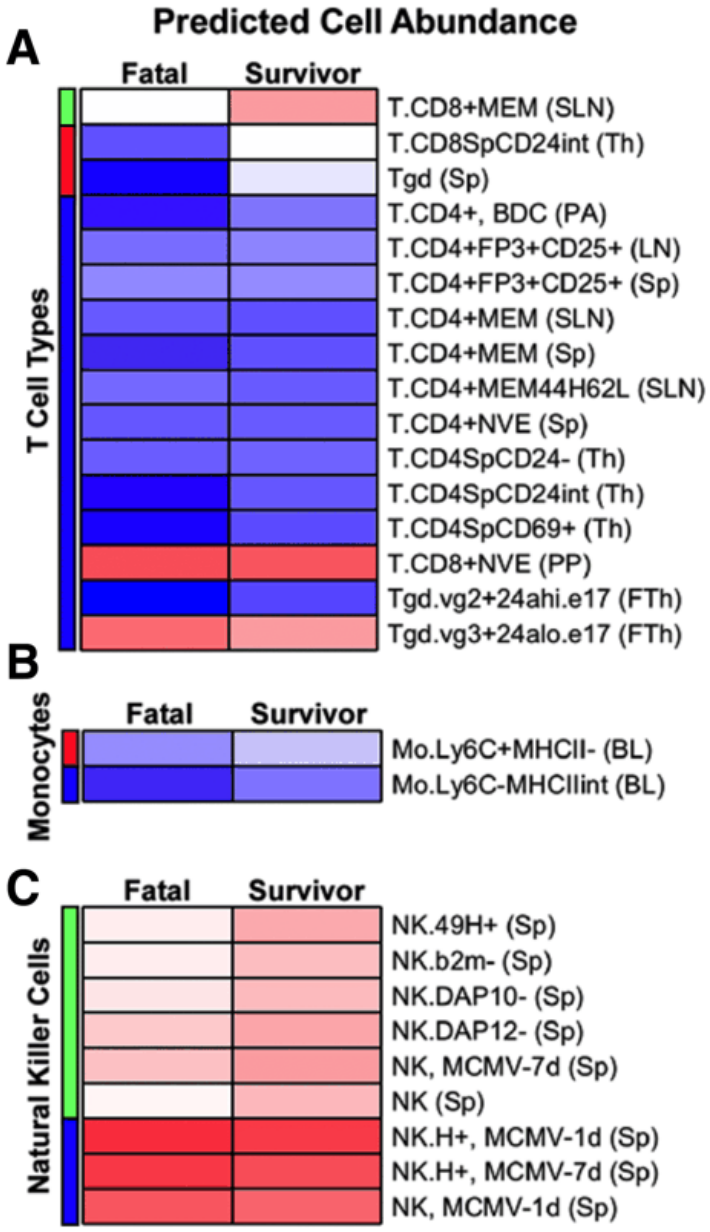


Figure 4.5: Predicted immune cell abundance in acute sufferers of EVD compared to convalescent controls. *Digital cell quantification performed and split into T cells (A), monocytes (B) and natural killer cells (C). As relative abundance to controls increases, the colour is s deeper red, as it decreases the colour is a deeper blue.*

4.3.4 The host response and the effect on patient outcome independent of viral load

It is readily accepted that viral load is a critical factor in patient survival, supported by the results of this study (Chapter 3, Figure 3.1) where viral load was significantly different between survivors and fatalities. Unsurprisingly, the above analyses involved patients with a diverse range of viral load. In order to confirm that the host response played an important role in patient outcome unconnected to viral load, an independent dataset was procured. This consisted of a group of 20 individuals (all testing negative for *Plasmodium spp.*) with an EBOV Ct value between 20 and 22, ten of whom died following hospitalisation whilst ten survived. This Ct range was chosen because it is not a reliable predictor of patient outcome, with a fatality rate of approximately 50%.

From the perspective of biological understanding of the disease, this allowed investigation into whether this differential expression of host factors was potentially independent of viral load. Upon mapping to the human genome and multidimensional scaling (MDS), these samples did not appear to cluster effectively according to their outcome (Figure 4.6). Differential gene expression analysis was performed using the transcriptomes of the samples in each group of patients. This identified the enrichment of very similar transcript abundance to that described in the original n = 141 analysis on the H_F and H_S patient groups. This suggests that the host response (at least at the transcript level in the blood) was mainly independent of viral load, with FGA, FGL1, ALB and FGB all being among the top ten differentially expressed genes (Table 4.2).

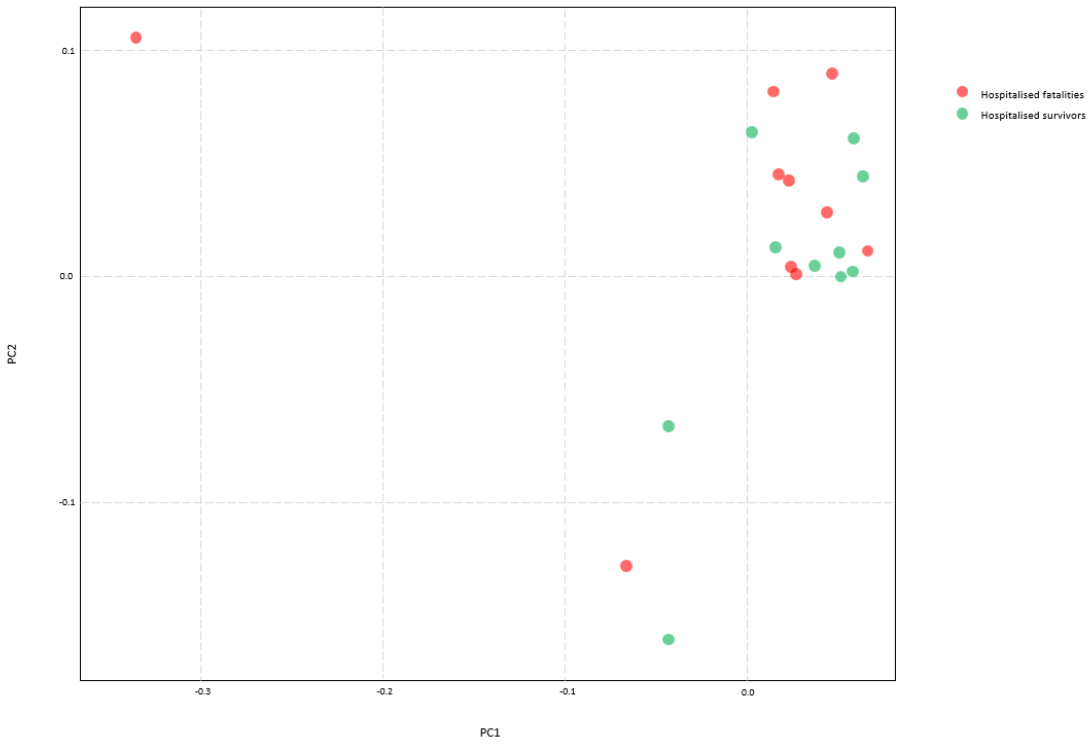


Figure 4.6: Principle component analysis (PCA) for ten hospitalised survivors (green) and fatalities (red) with comparable viral load. *The PCA was performed using the feature count values following alignment to the human genome and is based on the pairwise distances between samples.*

Table 4.2: Top ten differentially expressed genes between hospitalised survivors (n = 10) and fatalities (n = 10) with comparable viral load. *Those among the top ten DE genes in both the n = 20 and n = 138 studies are shown in bold.*

Gene name	Fold change H_F/H_S
Fibrinogen alpha chain (FGA)	2268.032377
C-reactive protein (CRP)	1325.288791
Fibrinogen like 1 (FGL1)	1315.405824
Albumin (ALB)	456.5393164
Fibrinogen beta chain (FGB)	450.5231655
Ceruloplasmin (CP)	349.1619036
Complement C9 (C9)	147.4703656
cAMP responsive element binding protein 3 like 3 (CREB3L3)	51.74968168
Prostaglandin reductase 1 (PTGR1)	43.94898822
Alpha-2-Macroglobulin (A2M)	38.44463466

4.3.5 The effect of patient age on host response

Although not identified in our dataset as being a significant factor (Chapter 3, Figure 3.3A), previous studies of EVD have shown age to be a determinant of survival (Sadek *et al.*, 1999; McElroy *et al.*, 2014a). In EVD outbreaks, individuals under the age of 5 and over the age of 45 appear to be at higher risk of fatality (Vega *et al.*, 2015; Li *et al.*, 2016). This is not an uncommon occurrence, with similar trends seen in HIV (Martin *et al.*, 1995; Darby *et al.*, 1996), CMV (Pawelec *et al.*, 2012) and Dengue (Guzman *et al.*, 2002), among others. It is possible that this association was not identified in this dataset due to the vast majority of patients being between the ages of 5 and 45. Despite this, investigations into which pathways are enriched in each age group was performed in order to potentially explain this phenomenon.

Patients were split according to their immunological age and outcome and KEGG analysis performed. A number of pathways were registered as being enriched in the 5-45 age group when compared to those under 4 and above 45, with a selection shown in Table 4.2. In the <4 group, the top pathways that are enriched in H_F compared to H_S are not observed to be differentially expressed. This makes sense as the haemostatic system is very different in infancy and adulthood (Andrew *et al.*, 1992). Therefore, the issues delineating fatal and non-fatal infections are different from those in the general population. Furthermore, although the pathways of complement, coagulation and cytokine interactions separated patients according to their outcome in individuals over 45, specific inflammatory responses, namely TNF- and NF-kB-signalling and chemokine production, were less important (Figure 4.7). This could be due to immunosenescence, with these pathways not functioning as well in an ageing population. In elderly humans, many facets of the immune response become

compromised, including the activity of neutrophils and NK cells (reviewed in Gomez *et al.*, 2008).

In the youngest age group (< 4 years), the malaria pathway was enriched in hospitalised fatalities, which gives support to this analysis as these individuals are at the highest risk of contracting the disease (Appendix 7.2). However, very few pathways were enriched in this age group. This is potentially due to the fact that it did not include many individuals. The absence of such immune response pathways must be considered with the small group size. In the eldest age group (> 45 years) some enriched pathways were associated with the inflammatory response, while others included those associated with cancer. This is unsurprising as this age group is at the greatest risk of such disease.

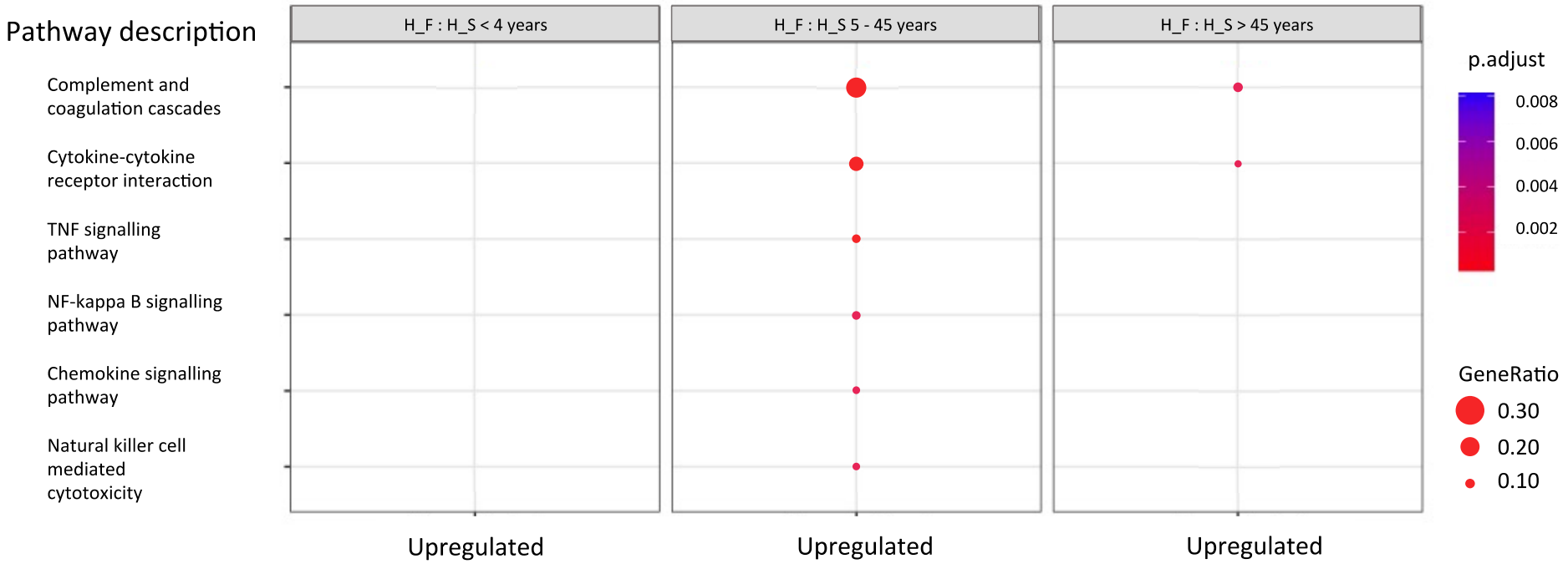


Figure 4.7: GO plot showing the KEGG pathways that are upregulated in hospitalised fatalities in each age group when compared to survivors. *The full tables can be found in Appendix 7.3.*

4.3.6 Prognostic capabilities of differentiated gene responses between patient groups

One application for differential gene expression analysis could be to identify suitable biomarkers for predicting the outcome of infection in individual patients. This could allow an improvement in triage, more effective allocation of treatment modalities and a logistical advantage to healthcare workers. This prediction would need to be achieved via the results of Nanopore sequencing or qRT-PCR in order to be completed rapidly enough to be beneficial.

As EBOV diagnosis is achieved via qRT-PCR, the most logical step appeared to be to pick a number of host genes and perform this semi-quantitative measurement on them. Here, VCAM1, ISG15 and TUBG1 were chosen due to their importance in the immune response and differential expression following the Illumina analysis (top 20 shown in Appendix 7.1). These qRT-PCR experiments were performed on a set of RNA extracts from whole blood of patients from the 2013-2016 EVD outbreak that were independent from those used in the previous analyses. This group comprised of 19 survivors, 20 fatalities and 6 uninfected controls (European individuals who would have never come into contact with the virus). The expression of VCAM1 and ISG15 were seen to be significantly higher in the H_F compared to H_S (Figure 4.8). There was not a significant difference between TUBG1 expression, however, this was included for completeness.

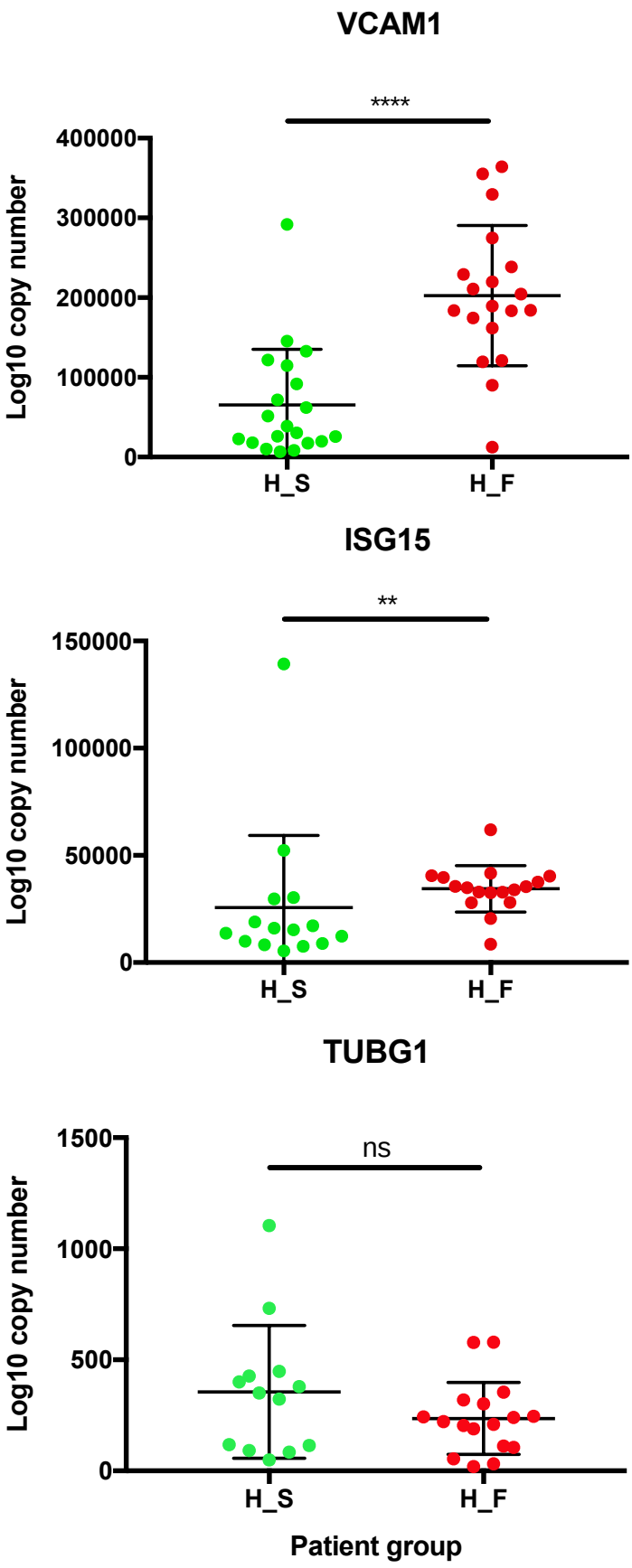


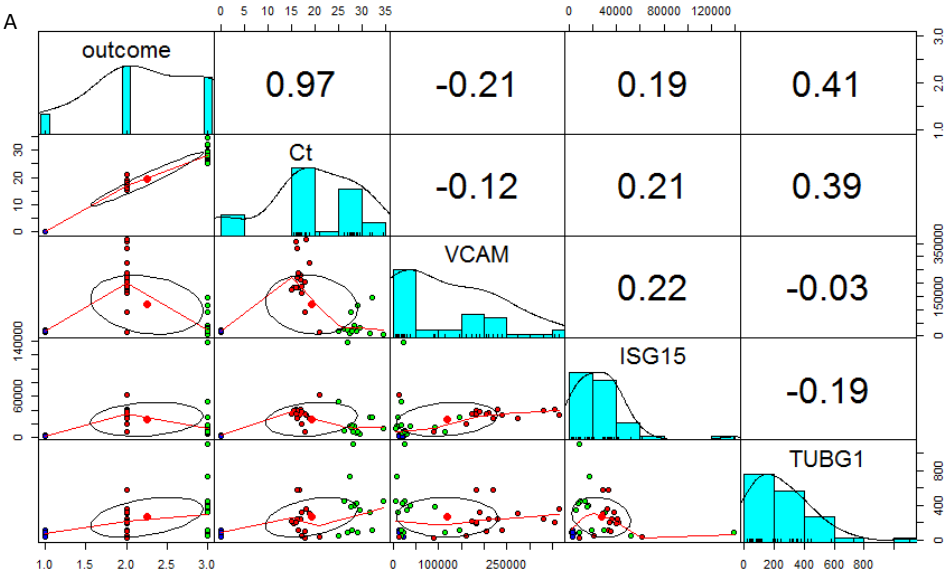
Figure 4.8: Differential abundance of target genes in an independent set of H_S and H_F following qRT-PCR. *Shown here as copy number mean + standard deviation, H_S*

*in green and H_F in red. Mann-Whitney test performed to compare groups. **** p<0.0001, ** p=0.0024, ns = not significantly different.*

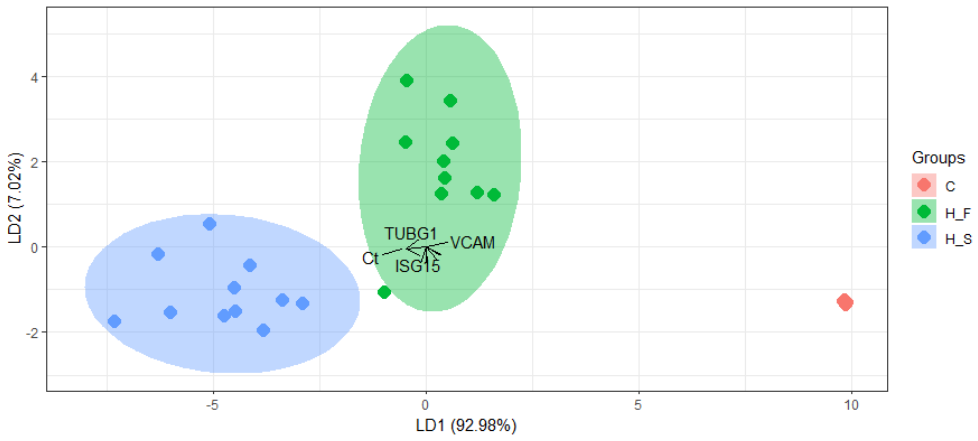
This allowed a linear discriminant analysis (LDA) to be performed, which takes quantitative information to differentiate binary groups. In this case, the differentiation of hospitalised survivors and fatalities using the calculated copy number of these selected genes as well as the viral load (Figure 4.9). This analysis calculates correlation coefficients for each factor and how they separate patients (Figure 4.9A). As expected, viral load is very effective in separating patient groups. This LDA was performed using a randomly selected training set of 60% of the tested samples (24 samples) and tested on the remaining 40% (15 samples). A linear model was able to successfully separate samples according to these factors (Figure 4.9B), and these were seen to be predicting survival with an accuracy of 100% even when EBOV Ct was used along with any single host gene classifier, despite this difference not being significant in TUBG1 according to the qRT-PCR. This is illustrated thoroughly in the separation achieved in the partition plot (Figure 4.9C). In this plot, each coloured area denotes a classification (in this case patient prognosis), thus the effective separation of these colours determines that using these two variables can effectively separate data into the correct groups. For example, the top left panel illustrates the effective separation achieved when using EBOV Ct and VCAM expression to predict patient outcome. However, when EBOV Ct is not included, this separation is much more difficult to achieve, as illustrated by the bottom right panel (Figure 4.9C). Here, TUBG1 and ISG15 expression cannot effectively separate samples into groups. These samples had EBOV Ct values that could be used to predict patient outcome, so the model was reconstructed without including EBOV Ct and an LDA performed again (Figure 4.10). Upon the development of this model, the training dataset was able to predict patient outcome with 83%, which increased to 90% using the testing dataset. This could suggest that using a multiplex PCR approach,

healthcare workers would be able to accurately predict patient outcome, even in cases with intermediate viral load.

A


outcome
Ct
VCAM
ISG15
TUBG1
0.97
-0.21
0.19
0.41
-0.12
0.21
0.39
0.22
-0.03
-0.19

B


LD2 (7.02%)
LD1 (92.98%)
TUBG1
VCAM
Ct
ISG15
Groups
C
H_F
H_S

C

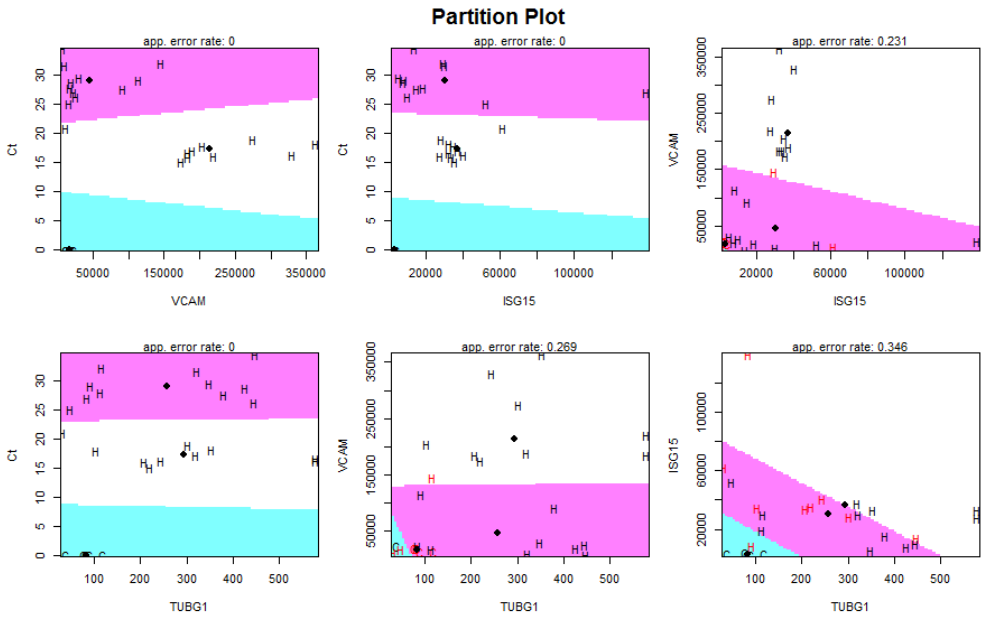
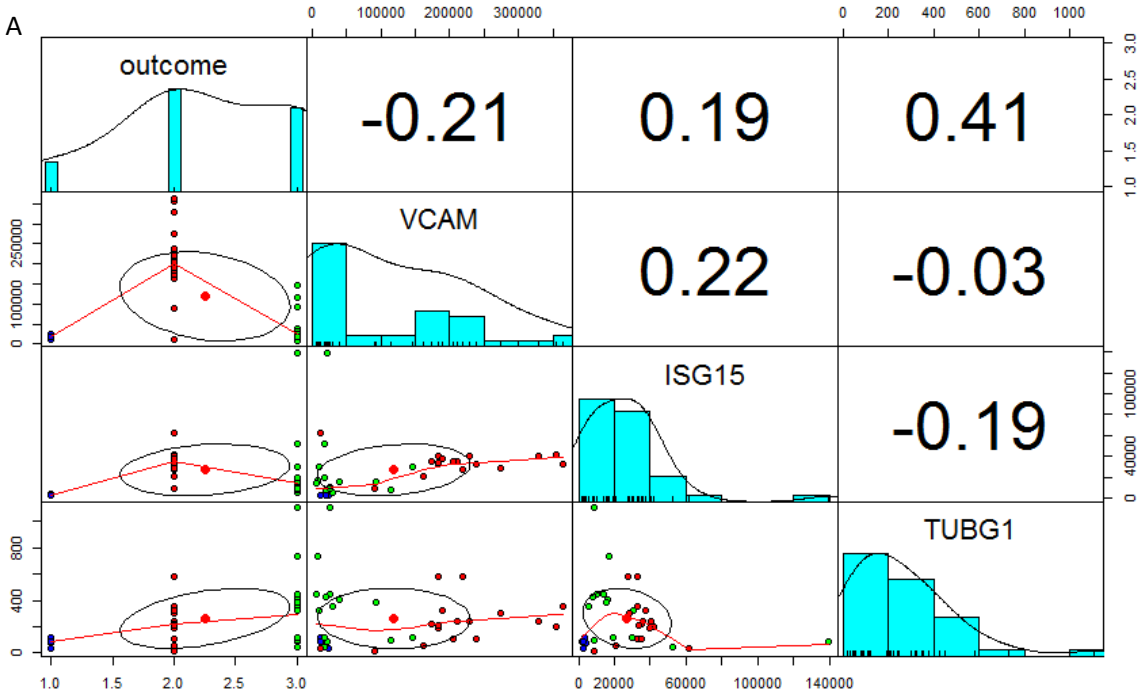

Partition Plot
app. error rate: 0
app. error rate: 0
app. error rate: 0.231
app. error rate: 0
app. error rate: 0.269
app. error rate: 0.346
Ct
VCAM
ISG15
TUBG1

Figure 4.9: Linear discriminant analysis comparing hospitalised survivors, hospitalised fatalities and controls. *Based on the copy numbers for three host genes: VCAM, ISG15, TUBG1 and Ct values for EBOV (denoted Ct). Correlation coefficients following linear discriminant analysis (A), bi-plot (B) and partition plot (C).*

A



B

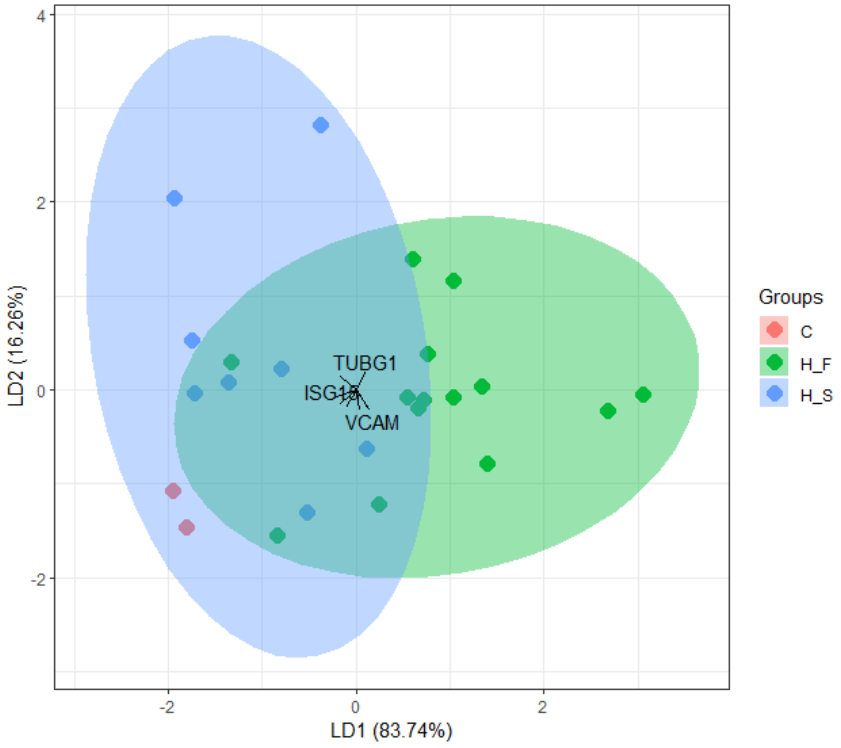


Figure 4.10: Linear discriminant analysis in the absence of viral load. *Based on the copy number for three host genes: VCAM, ISG15, TUBG1. Correlation coefficients following analysis (A) and bi-plot (B).*

4.3.7 Using Nanopore sequencing to study the host response

Illumina sequencing of our samples provided a substantial quantity of data and many interesting avenues to study. The development of a prognostic tool creates the potential for the influence of patient management. Patients with similar levels of care could be housed together, thus improving the efficiency of the healthcare workers. A challenge associated with Illumina sequencing is that the analysis takes several weeks and can only be performed by individuals who have undergone extensive training. Following this period, acutely ill individuals would no longer be in the ETC. Nanopore sequencing, which was employed in the field at EMLab sites for phylogenetic purposes, can be performed rapidly and with minimal training. Long reads are accessible within just 15 minutes. This provides the potential for real-time sequencing and mapping to the human genome, with the results being potentially used to inform patient treatment (for example the provision of αIFNs etc.) or to allocate limited resources to patients.

Six RNA samples extracted from the whole blood of three survivors and three fatalities, were sequenced using a MinION (Nanopore Technologies, Oxford) and the reads mapped to the human genome. The samples did appear to cluster successfully according to patient outcome following PCA analysis (Figure 4.11A). Despite this, the number of passed reads collected from Nanopore sequencing was much lower than that using Illumina sequencing. However, many of the same genes were upregulated in the fatalities compared to the survivors as shown by the Illumina data (Figure 4.11B).

Despite these similarities in general trends, the top DE genes following Nanopore sequencing were not necessarily as expected. Genes that had not previously been

highlighted as being associated with EVD within the aforementioned Illumina study or elsewhere were identified. However, at a systems biology level there did appear to be an enrichment of particular pathways associated with the immune response. Additionally, a number of specific genes with an overwhelming pro-inflammatory response are seen to be upregulated in H_F compared to H_S. This included ISG15, which has been shown to have predictive ability in EVD patients.

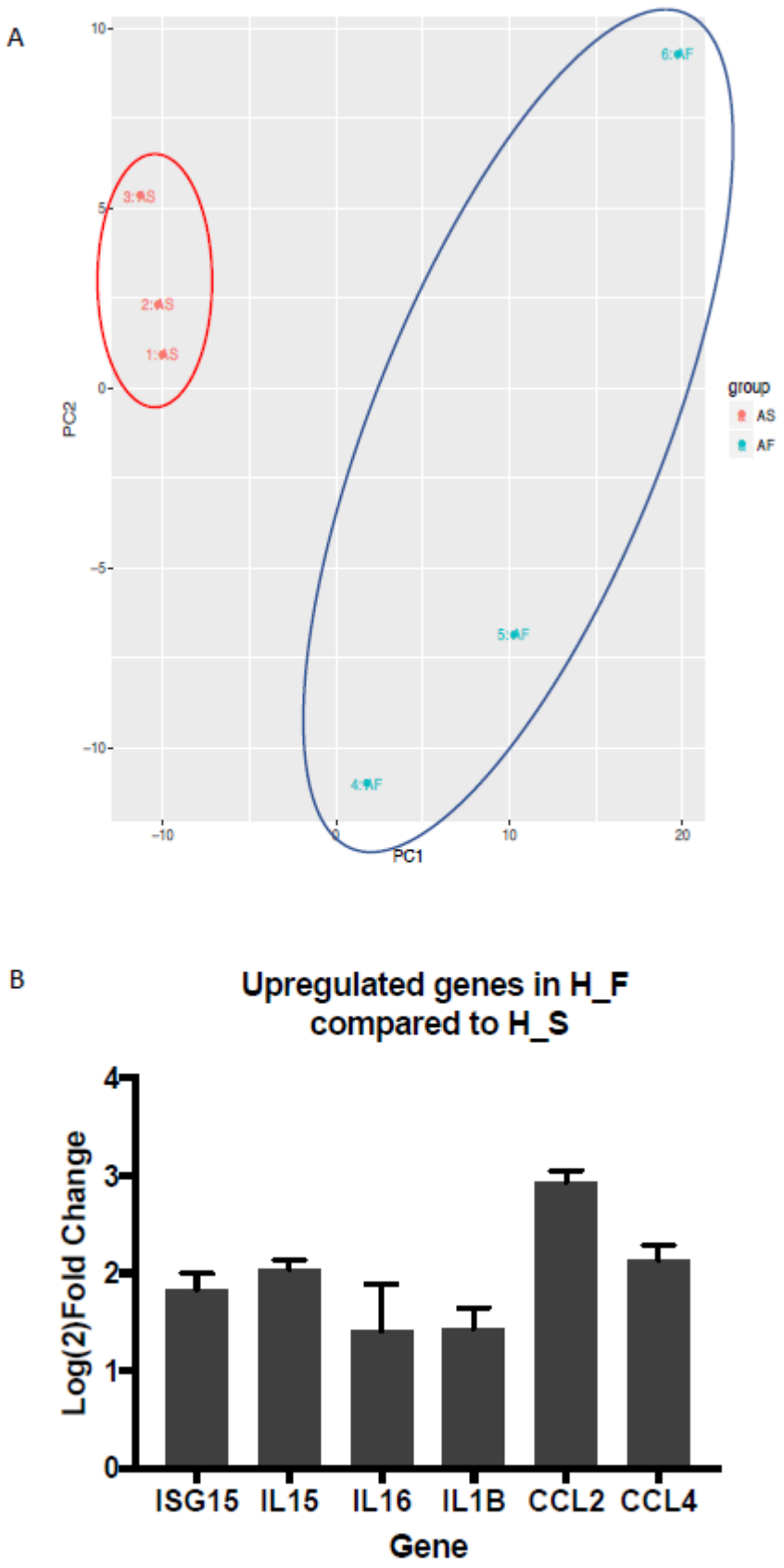


Figure 4.11: Differentially expressed genes in fatalities compared to survivors of EVD following Nanopore sequencing. *Following mapping to the human genome, a PCA was performed to show grouping of samples (A), with six genes associated with a pro-*

inflammatory response shown to upregulated in fatalities compared to survivors (B). Three H_F and three H_S samples were sequenced on a MinION platform and their levels of gene expression compared using DESeq2.

4.3.8 A comparison of experimentally infected samples to naturally infected humans

Previous studies have highlighted the similarity between the response of humans and NHPs to EVD (Geisbert *et al.*, 2000; Hensley *et al.*, 2012). However, due to the cost associated with working with NHPs, other mammalian models have been developed. One such model is that of the guinea pig. Guinea pigs are initially asymptomatic to EBOV infection, but this alters as the virus adapts to the host upon sequential passage (Volchkov *et al.*, 2000; Dowall *et al.*, 2014). Following adaptation, EBOV becomes more virulent and produces fatal infections coupled with an extensive immune response with many similarities to that seen in humans (Connolly *et al.*, 1999; Padilla-Carlin *et al.*, 2008; Nakayama and Saijo, 2013).

Previously, colleagues had performed infection experiments with this model organism in containment level (CL) 4 conditions at Porton Down (Public Health England) (Dowall *et al.*, 2014). Here, RNA was extracted from the pooled spleens of six infected guinea pigs. This allowed a comparison between these experimentally infected guinea pigs and naturally infected humans from the West African outbreak. Though there is a difference in terms of the source material, general trends in cytokine activity could still be observed.

Dowall *et al.* (2014) infected guinea pigs with EBOV and used a homogenate prepared from the spleens of infected subjects to infect the subsequent group. This continued until an increase in pathogenicity was demonstrated. At each stage, the initial viral concentration was kept constant. From these passages, RNA was extracted and sequenced using a MiSeq Illumina platform. This allowed the analysis of both the host response as well as the virus itself. Following transcriptomic analysis, a number of

critical immune response markers were seen to elevate across this time (Figure 4.12). These include a number of genes associated with interferon (IFN) signalling.

IFN stimulated gene (ISG)15 and IFN-induced protein with tetratricopeptide repeats (IFIT)5 have been previously shown to be significantly upregulated in acute EVD patients in comparison to a convalescent control group (Caballero *et al.*, 2016; Liu *et al.*, 2017). In mouse models, the upregulation of ISG15 has been observed to have protective qualities (Malakhova and Zhang, 2008; Okumura *et al.*, 2008). However, in the guinea pig experimental model the upregulation of ISG15 increases through sequential passages and along with the fatality rate. IL-18 is an inducing factor of IFNγ and has been previously found to be upregulated in infected monkeys (Hensley *et al.*, 2012). Together, these data suggest that IFN signalling plays a critical role in EBOV pathogenicity in guinea pigs as well as other mammals.

A number of cytokines and chemokines were also observed to be expressed in greater abundance in later passages. This follows the pattern observed in the human data as the abundance of pro-inflammatory cytokines increases alongside severity of infection. The abundance of these cytokines can lead to a cytokine storm and is associated with severe immunopathology (Zampieri *et al.*, 2007). For instance, CCL2 is a chemokine capable of recruiting immune cells such as monocytes and dendritic cells (Conrady *et al.*, 2013). CCL2 and CXCL10 have been implicated as having a key role in pulmonary immunopathology following respiratory syncytial virus (RSV) infection (Culley *et al.*, 2006). Tumor necrosis factor (TNF) makes up an important part of the acute phase response, a pathway also highlighted to be enriched in acute EVD sufferers and associated fatal infections (Baize *et al.*, 1999; Hensley *et al.*, 2012; Liu *et al.*, 2017).

CXCL11 is a recruiter of activated T lymphocytes, which has been shown to be upregulated in EBOV-infected macaques (Caballero *et al.*, 2016). In the guinea pig model, there were distinct differences in transcriptomic profiles between passages, with a number of genes being differentially expressed between passage 2 and passage 5. Passage 1 was not used as a baseline as the initial injection likely had a substantial effect on the immune response. This potentially explains the pattern observed in Figure 4.12.

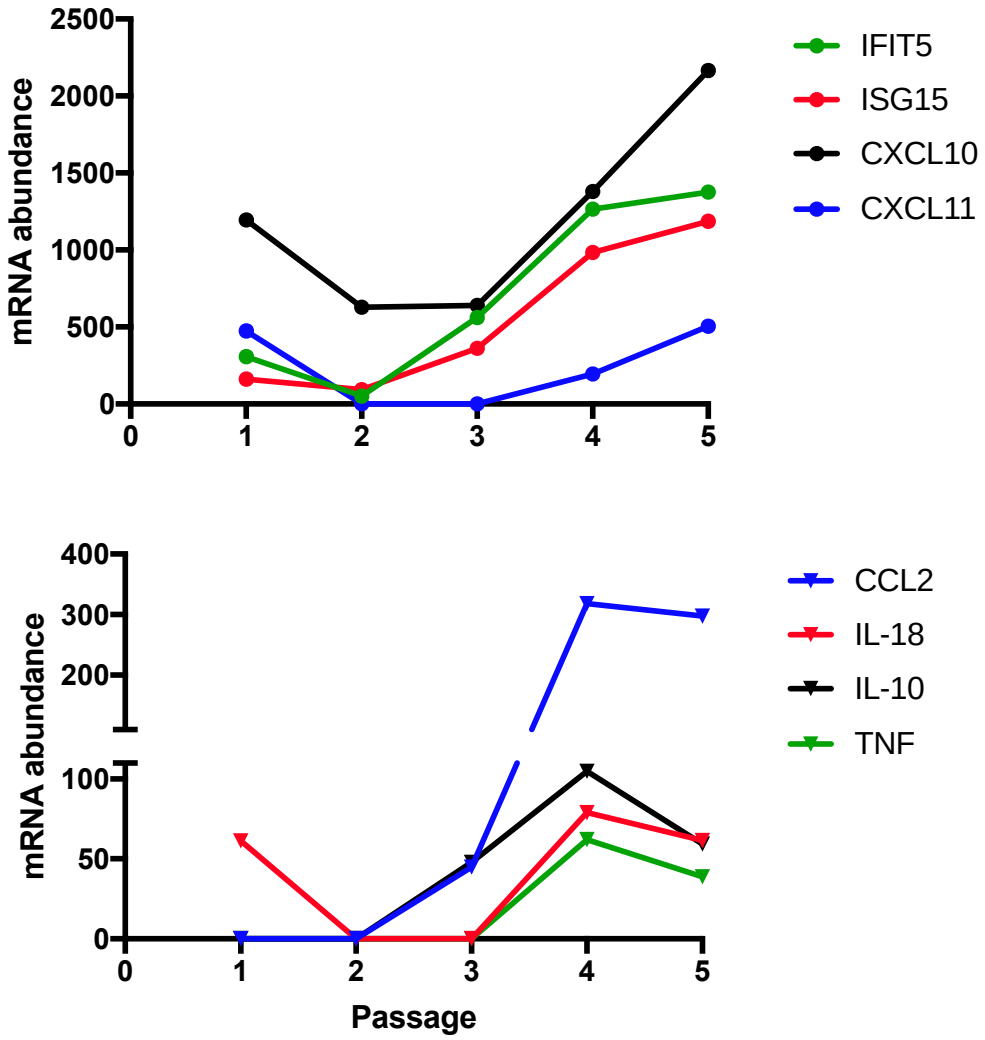


Figure 4.12: Gene expression levels of immune response markers across passages. *Each data point shows the value from sequencing a pooled sample from four guinea pigs in each passage. Read depth represents the fragments per kilobase of transcript per million mapped reads (FKPM).*

The top DE genes (Figure 4.13) show a considerable change in mRNA abundance, showing the significant effect that EBOV infection imparted on subjects. These genes have all been demonstrated to play significant roles in viral infections. Previous experimental data suggested that viral protein VP24 was able to inhibit IFN signalling via the blocking of STAT1 (Reid *et al.*, 2006; Zhang *et al.*, 2012), however this is not reflected in experimentally infected animals or natural infections (Caballero *et al.*, 2016; Liu *et al.*, 2017). Here, STAT1 is highly upregulated in later passages. HSPA8 expression is upregulated in cases of cellular stress including infection and inflammation as exhibited by Tsuboi *et al.* (1994), including in response to Epstein-Barr virus (HHV4) (Stricher *et al.*, 2013). RIPK2 appears to play a key role in a number of viral infections, including apoptosis in dengue infection (Morchang *et al.*, 2011) and the dampening of immunopathology in influenza via IL-18 signalling (Lupfer *et al.*, 2013).

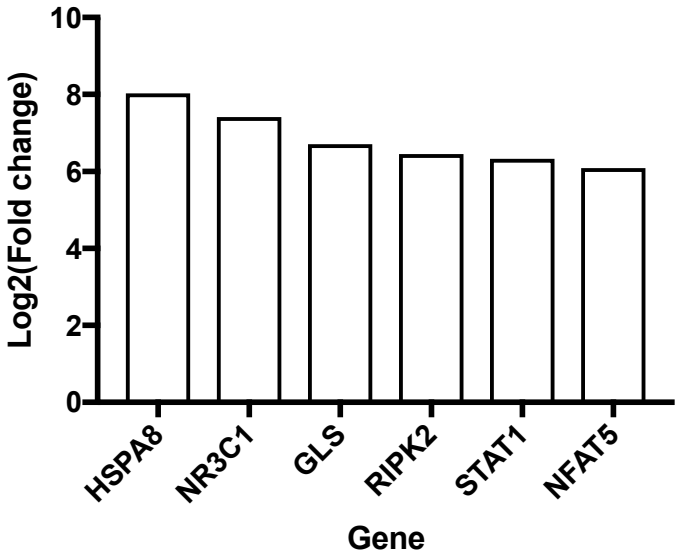


Figure 4.13: Top differentially expressed genes from passage 2 to passage 5 in EBOV-infected guinea pigs. *Mean fold change (log2) is shown as a proportion of FKPM expression values. No error bars are shown as samples were pooled before sequencing.*

Once again, IPA was employed to identify certain pathways associated with the final passage. These data predicted the enrichment of specific signalling pathways associated with the immune response (Figure 4.14). NF-kB has often been implicated in viral infections due to its role in immunity and apoptosis. NF-kB has been associated with IL-18 and NR3C1 expression in hepatitis C infections (Capone *et al.*, 2014). IL-8 signalling was also identified to be increased in later passages, as well as in fatal human infections (Liu *et al.*, 2017). This enrichment of neutrophil and granulocyte chemotaxis could be responsible for much of the immunopathology experienced by the guinea pigs in this study. A higher proportion of subjects in later passages died in comparison to earlier ones, so this finding parallels what has been seen in outbreaks.

JAK/STAT signalling was highlighted as being enriched in passage 5, which is not surprising as STAT1 was one of the top differentially expressed genes in these samples (Figure 4.13). The activity of IFN can trigger the JAK/STAT signalling cascade, which in turn leads to the transcription of a number of ISGs (as reviewed by Fensterl and Sen, 2009). Of the top differentially enriched pathways, the vast majority were upregulated later in infection. This phenomenon is readily observed in naturally infected individuals.

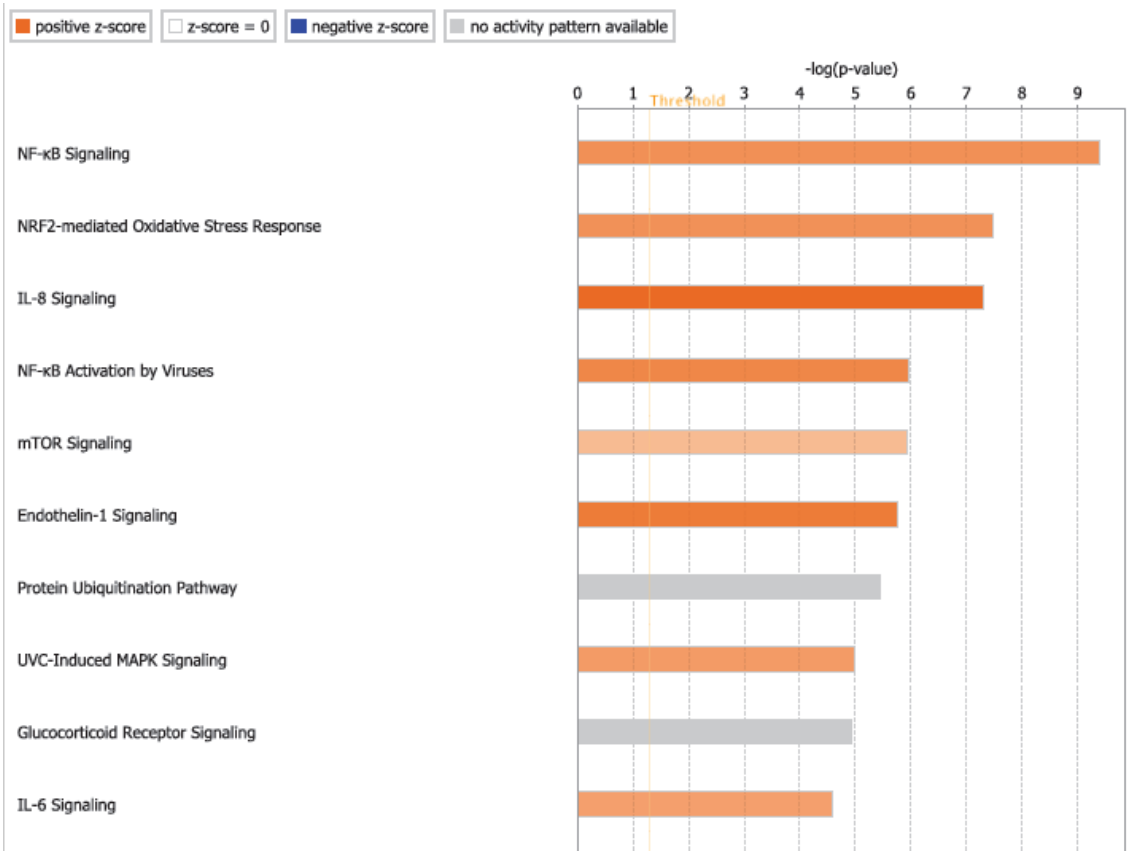


Figure 4.14: Top canonical pathways differentially enriched from passage 2 to passage 5 in EBOV-infected guinea pigs. *Analysis performed using IPA (Qiagen).*

Following the pathway analysis, a DCQ was performed using ImmQuant, due to the analysis of samples from rodents (Frishberg *et al.*, 2016). This enabled a comparison to be made with the one performed using human data (Figure 4.5) and allowed an immune cell profile to be predicted. This was performed as cellular abundance can have a dramatic influence on the transcriptomic profile. Here, a number of key immune cell types can be seen to increase in response to infection (Figure 4.15). Notably, there is increase in neutrophils across passages, which supports the detection of the IL-8 pathway as being enriched in Figure 4.14. Additionally, the gradual increase of mature CD56+CD16+CD3- NK cells shows an attempt to control infection. NK cells expressing CD56 and CD16 usually account for the majority of circulating NK cells and have cytotoxic functions, being significant producers of a number of proinflammatory cytokines (Cooper *et al.*, 2001). NK cells have been implicated as playing an important role in EBOV infection in humans (Cimini *et al.*, 2017; Liu *et al.*, 2017).

The steady increase in the presence of eosinophils was likely due to the formation of eosinophilic intracytoplasmic inclusions, which have been noted previously in livers of EBOV-infected cynomolgus monkeys (Geisbert, *et al.*, 2003). These inclusion bodies represent sites of viral replication, formed from aggregates of viral proteins. The increase of eosinophils alongside that of viral load is commonly observed in a number of other viral infections including rabies (Lahaye *et al.*, 2009).

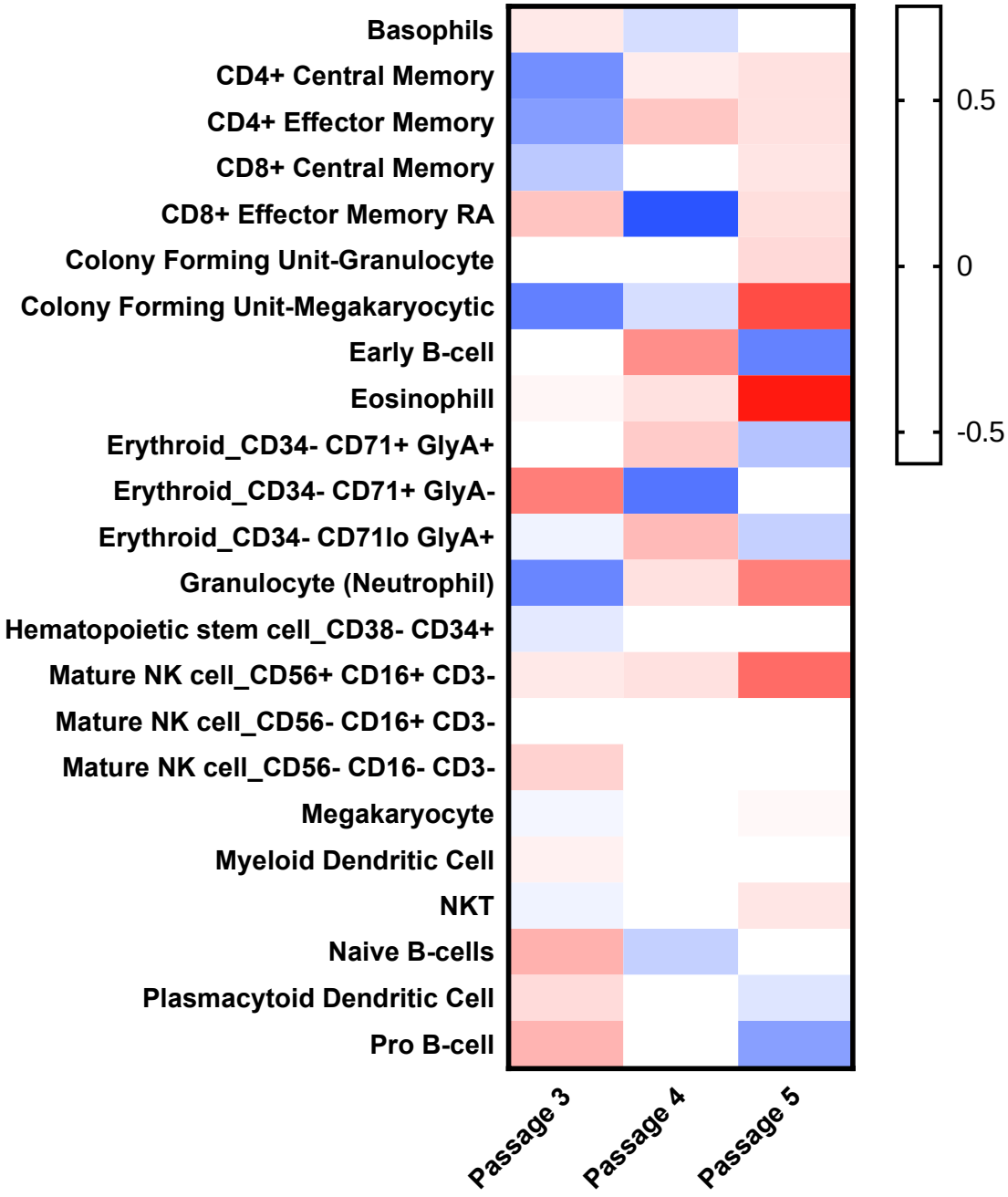


Figure 4.15: Predicted immune cell profile of passages 3, 4 and 5 in comparison to the cell abundance of passage 2. *Heatmap constructed following digital cell quantification, with the colour legend representing the relative abundance (log2) of each cell type compared to passage 2.*

4.4 Discussion

Studying EVD carries a number of challenges, not only due to its requirement for high-containment laboratories, but also due to the fact that it affects countries in low-resource settings, negatively impacting the ability of healthcare workers to effectively monitor patients and provide care. There are debates as to the best way to study EBOV and its effects. Several appropriate animal models exist; however, these do not always mirror the human course of infection (Safronetz *et al.*, 2013). In this study, diagnostic testing for EBOV has provided RNA extracts from the whole blood of infected individuals and allowed investigation into the host response to infection. These findings highlight the insights that transcriptomic analysis can provide into improving clinical management of disease.

The ISG pathway was observed to be heavily enriched in this cohort of EVD patients compared to controls. This finding is noteworthy due to previous associations of the ability of EBOV to dampen the IFN response (Basler *et al.*, 2000; Basler and Amarasinghe, 2009). More recently, IFN β responses have been found in expatriated EVD patients with moderate disease (McElroy *et al.*, 2016), which studies in NHPs have postulated as potentially providing some protection against EBOV and Marburg virus (MARV) (Smith *et al.*, 2013). In our Guinean cohort, the substantial increase of IFN activity in severe, eventually fatal infections may suggest the opposite. Certainly, a link between the activation of IFN-like pathways and survival cannot be established in the absence of additional palliative care as these patients tended to have a poorer prognosis. This is mirrored in NHPs given a lethal challenge of EBOV, where an

uncontrolled IFN response may decrease an individual's chance of survival (Rubins *et al.*, 2007). Potential mechanisms for this include the formation of an uncontrolled cytokine storm as well as lymphocyte apoptosis (Hensley *et al.*, 2012). IFN signalling has been linked to immunopathology in many viral infections, with the uncontrolled production of pro-inflammatory cytokines often correlating with disease severity. Influenza and respiratory syncytial virus are examples of this (Ostler *et al.*, 2002; Jong *et al.*, 2006). This cytokine-mediated immunopathology is readily reproducible in animal models, and these have been critical in furthering our knowledge of EBOV pathogenesis. Here, a robust immune response was observed in guinea pigs following EBOV infection and the adaptation of the virus to the host. This mirrored the human response well in terms of transcriptomic analysis of samples from the 2013-2016 outbreak in West Africa. This is also reflected at an immune pathway level, with IFN-associated pathways and those involved in inflammation being upregulated in later passages, when guinea pigs were experiencing higher fatality rates and increased viral load. DCQ analysis indicated an upregulation of NK cells and eosinophils, suggesting the importance of cytotoxic immune cells as well as the formation of eosinophilic intracytoplasmic inclusions in the spleens of infected guinea pigs. However, this is based on the sequencing of the pooled spleens of infected guinea pigs and so caution should be taken when comparing these samples. Furthermore, the pooling of these organs means that only the overall trend can be investigated. Therefore, some of the individual effects observed in each organism may have been masked.

Another majorly enriched pathway was that of complement and the coagulation cascade, also found following the KEGG analysis, with fibrin deposition having been

previously associated with EBOV infection (Geisbert *et al.*, 2003; Feldmann and Geisbert, 2011). However, much of the mRNAs associated with such genes are not readily found in the blood and considered to be liver-specific. In this outbreak, fewer individuals with EVD were observed to be suffering from haemorrhagic symptoms (Chertow *et al.*, 2014), but liver and kidney damage was very common in the later stages of disease. This aberrant organ function was confirmed by looking at levels of aspartate transaminase (AST) and alanine transaminase (ALT) (Lyon *et al.*, 2014; Schieffelin *et al.*, 2014). Therefore, the identification of these mRNAs could be due to leakage of hepatic mRNAs into the blood. This would explain why these transcripts were more abundant in the blood of those who subsequently succumbed to infection.

One of the limitations of this study is that any conclusions made can only be applied to the blood transcriptome of individuals with EVD. Therefore, a critical finding of this study was the similarities to protein abundance in cytokine and chemokine expression, both in this outbreak (Ruibal *et al.*, 2016) and several previous in Gabon and Republic of Congo (Wauquier *et al.*, 2010). This is a particularly pertinent consideration in analysing outputs of certain tools employed in this study, for example the DCQ. This presents a prediction of the immune cell profile based on gene expression data. However, many of the genes this is based on can be influenced by other factors and therefore should be interpreted carefully. One such example is the predicted increase in NK cells due to the presence of IFNγ (Figure 4.5). Although a known marker for NK cell activity, this cytokine can also demonstrate the activity of other cell types and associated pathways.

Here, major differences are observed between the transcriptome taken from hospitalised survivors when compared to their fatal counterparts. The explanation for this could be due to the differential activation of gene transcription but also the potential difference in immune cell profiles between these groups. Cytokines and chemokines lead to cell recruitment, so the difference in immune response will directly affect the infiltration of specific cell types. Such differences were predicted using a DCQ analysis, which noted a higher proportion of monocytes in hospitalised survivors. This finding was validated using FACS on similar patient samples (Liu *et al.*, 2017), providing confidence in the accuracy of the DCQ. NK cells were highlighted as also being more abundant in hospitalised survivors, which is unsurprising as these cells have been suggested to provide a protective effect in EVD in experimentally infected mice (Williams *et al.*, 2015). This priming of the innate immune system may therefore be critical in successfully fighting an EBOV infection.

These differences allowed certain genes to be selected for further qRT-PCR and an LDA performed by taking this quantitative information and using it to separate H_S and H_F. In this case, it was used to effectively predict patient outcome and therefore has the potential to be included in the formation of a prognostic tool in the form of a multiplex PCR. During the construction of this LDA, age was initially included in the analysis, however, this had no beneficial impact and was therefore removed. qRT-PCR is already performed to confirm the presence of EBOV, so the development of an effective multiplex PCR would not delay any provision of diagnosis. When EBOV Ct was included in the LDA, even with a sample size of only 20, this information, along with one or more host genes, allowed patient outcome to be correctly predicted in every

case. Even when EBOV Ct was not included in the LDA, patient outcome was able to be predicted with an accuracy of 90% in the testing dataset, representing 40% of the independent samples. This could provide useful information to healthcare professionals and inform clinical management of patients, which is particularly important in low-resource settings. This method could be particularly valuable in cases where viral load is not an effective indicator of patient outcome, including herpes simplex encephalitis (HSE), caused by herpes simplex virus (HSV) (Saraya *et al.*, 2016). HSE is particularly notable as it is responsible for a large proportion of the global cases of viral encephalitis and has high rates of mortality (Solomon *et al.*, 2012).

In a future outbreak, the accuracy of this model is likely to increase over time as increasing data points are added. However, it should be noted that it is possible that the success of this model may not extrapolate to further samples. Testing on a blinded sample set would be an effective test of the model, where the outcome of the patients is unknown. In these cases, patient information such as age and malaria status could also be included to aid in the accuracy of outcome prediction.

The other option for the formation of a prognostic tool would be to use Nanopore sequencing at the ETC. This real-time sequencing method was performed for phylogenetic purposes during the 2013-2016 outbreak, showing the ease of implementing such methodology, even in such low-resource settings (Quick *et al.*, 2016). However, in this study, a transcriptomic approach was taken to study the RNA extracts from EVD patients. Successful reads were mapped to the human genome using BWA-MEM. Here, the quality of the RNA sample was found to be extremely

important in the success of each MinION run. Lower quality RNA extracts yielded substantially fewer reads than those of higher quality. This was much less dramatic using Illumina as a sequencing method. The RNA used with this method had been stored for over a year at -80°C, with few freeze-thaw events. Although the conditions were kept optimal for RNA storage, some degradation is inevitable (Auer *et al.*, 2014). It was found that upon sequencing of fresh extracts from clinical samples, using Nanopore sequencing with the same kit produced increased read length and quantity. Additionally, this approach included a PCR step, introducing a potential bias towards shorter transcripts.

However, Nanopore platforms have been used effectively to sequence clinical samples, thus could be implemented for use in clinical settings if the process were to be refined. The general trends observed between H_F and H_S were found to be very similar when comparing analysis of Illumina and Nanopore data. For the moment, Nanopore sequencing represents a significantly more expensive method of determining patient prognosis, therefore supporting the use of qRT-PCR of targeted genes. The identification of enriched pathways in the hospitalised survivors at different immunological ages could also inform clinical practice and be a step forward in personalised treatment of EVD. The differential enrichment of certain pathways in different age groups also points towards a benefit to a more personalised approach to treating EVD.

Delving further into the genetics of patients with EVD could also delineate some of the factors influencing patient prognosis. This is challenging given the transcriptomic data acquired in this study. However, one option would be to use software capable of

predicting HLA types from such data in order to compare patient groups. Additionally, a SNP analysis of the human genome of this dataset could help to better understand the host response.

Chapter 5: The impact of co-infections on patient outcome during Ebola infection

5.1 Introduction

5.1.1 Activation of the interferon response during EVD

The 2013-2016 outbreak in West Africa has provided many insights into EBOV pathology and EVD. The utilisation of next-generation sequencing (NGS) during this outbreak has allowed investigation into the virus itself as well as the antiviral responses of hosts. One observation that has been highlighted is the difference between the host response that we observe in experimental *in vitro* infections and the response of patients, particularly in the acute immune response. One example of this is the role that interferons (IFNs) and interferon-stimulated genes (ISGs) play in the host response. These were long thought to be successfully supressed by viral genes VP35 and VP24 (Basler *et al.*, 2000; Basler and Amarasinghe, 2009), however; this is not supported by studies of infected patients (Caballero *et al.*, 2016; Liu *et al.*, 2017). Real-life infections represent a much more complex system than that of *in vitro* infections. One stark difference is the underlying burden of disease that we observe in many of the countries where Ebola outbreaks have occurred. This could be a pivotal factor in the immune response and why we don't see this IFN response during experimental infections. IFNs are cytokines grouped into three classes: types I, II and

III. Type I IFNs (most importantly IFNα and β) are secreted following detection of viral particles. IFNγ, belonging to type II, is secreted by T lymphocytes and natural killer (NK) cells following their activation. These IFNs act to create an antiviral state, playing a key role in the host immune response. Type III IFNs can also support the formation of an antiviral state, but their distribution is more limited due to production by restricted cell types (Randall and Goodbourn, 2008).

5.1.2 Plasmodium falciparum as a co-infection

In Guinea, where the majority of our samples originate, the incidence of malaria is high (367.8 cases per 1000 population in 2017) and 100% of the population are at risk of exposure (WHO, 2018). The protozoan parasite species responsible for these cases is *Plasmodium falciparum*. Guinea is considered to be one of the poorest countries in West Africa, with only half of the population having access to public healthcare services. This can act as a barometer for how substantial the burden of disease could be among the population. Deaths due to malaria and lower respiratory infections are responsible for almost a quarter of all deaths in Guinea (12.5% and 10% respectively) and malaria is the biggest killer of those under the age of five (WHO, 2018). Malaria as a co-infection with EVD has been implicated in influencing patient outcome (O'Shea *et al.*, 2016; Rosenke *et al.*, 2016). However, the direction that this takes is not transparent.

5.1.3 Viral and bacterial co-infections

There is a plethora of other co-infections that have been observed to impact patient outcome. Lauck *et al.* (2014) found that co-infection of *Zaire ebolavirus* (EBOV) with GB

virus C (GBV-C, formerly hepatitis G virus) was associated with patient survival. Human herpesvirus 4 (HHV4, formerly Epstein-Barr Virus) is a member of the Herpesvirus family and has a high incidence in developing countries (Chakraborty *et al.*, 2010; González Saldaña *et al.*, 2012). In a healthy individual this virus can coexist with the host, but as an opportunistic infection it can have deleterious effects. Reactivation of the virus has been observed in individuals with EVD following the impairment of the host immune response (Agrati *et al.*, 2016). Throughout the course of EVD gastrointestinal involvement is very common. This can lead to the translocation of pathogens from one body compartment to another (Kreuels *et al.*, 2014; Wolf *et al.*, 2015). Incidences of bacterial septicaemia have been noted in cases of EVD (Kreuels *et al.*, 2014), but not fully investigated due to the nature of an outbreak situation of a viral agent. Suspected bacterial infections are usually treated following clinical rather than laboratory diagnosis (Hunt *et al.*, 2015).

5.1.4 The effect of age on co-infections

The link between patient age and their probability of surviving EVD has been documented previously, with those who are very young or very old experiencing higher fatality rates (Sadek *et al.*, 1999; WHO, 2014). Older EVD patients are at greater risk due to immunosenescence, defined as the exhaustion of the immune response. Persistent infections cause the dysregulation of many parts of the immune response, with notable differences occurring in the T cell compartment as well as the innate response (as reviewed by Fülöp, Larbi, & Pawelec, 2013 and Solana et al., 2012 respectively). This effect is particularly notable in populations experiencing a substantial burden of disease. Due to this mechanism, this effect is also observed in

other viral and bacterial infections. Additionally, the presence of certain co-infections would be expected to be more prominent in certain age groups. This is important as it further supports the functionality and validity of our analysis pipeline.

5.1.5 Identification of potential co-infections

In this study, Illumina RNA-sequencing was used to identify the presence of microorganism-specific nucleic acids in RNA extracts of whole blood representing potential co-infections. This has been performed in previous studies using target independent sequencing techniques for respiratory and haemorrhagic viruses as well as other potential pathogens (Fischer *et al.*, 2015; Thorburn *et al.*, 2015; Dinçer *et al.*, 2017; Ramesh *et al.*, 2018). In a study of viral respiratory pathogens, a cut-off Ct value of 32 was posited as providing enough information for viral identification by next-generation sequencing (Thorburn *et al.*, 2015). This unbiased approach allows the identification of viral, bacterial, fungal and parasitic sequences in diagnostic samples.

At the Ebola Treatment Centre (ETC), *Plasmodium spp.* were tested for via an RDT (BinaxNOW Malaria, Alere, ME). The RDT allowed the identification of cases of *P. falciparum* and those individuals who could be infected with a different or multiple species of the parasite. The T1 band (Figure 5.1) detects the presence of histidine-rich protein 2 (HRP2) – a protein specific for *P. falciparum*. The T2 band detects *Plasmodium spp*. aldolase, found in the four species of malaria capable of infecting humans. These different techniques provided an interesting opportunity for comparison between methodologies of *Plasmodium spp.* detection.

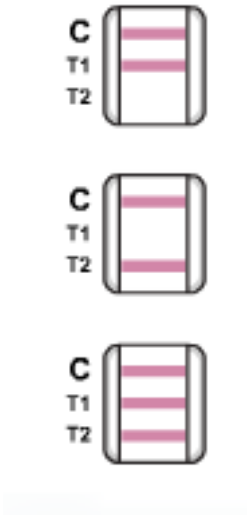


Figure 5.1: Outcome following a BinaxNOW Malaria test. *The top panel represents a positive P. falciparum infection, below that a Plasmodium infection that is not P. falciparum, then a mixed P. spp. infection.*

5.1.6 Geographic location of infections and the effect of forest area on fatality rate

The samples from this study are taken from Guinea (in the majority), Liberia and Sierra Leone. These West African countries are covered, in part, by forest. This can increase the rates of transmission due to contact with zoonotic reservoirs and also access to potentially infected bush meat. Additionally, those further from ETCs and access to clinical attention may experience higher fatality rates. On the other hand, living in urban environments poses its own dangers, as people are housed in a more confined environment and therefore human-to-human transmission becomes more likely.

5.1.7 Objectives and hypotheses

The hypothesis was that the presence of co-infections could impact the host immune response, which could be observed at the level of gene expression. Moreover, these co-infections could be the driving force behind some of the IFN expression we observe in genuine infections but not experimental ones. Furthermore, bacterial species found in blood samples represent translocation of species from the gut and therefore significant damage to the gut wall. Thus, the abundance of these could be higher in those who go on to succumb to infection, in comparison to those who survive.

The objective was to identify the co-infections present in blood samples from patients during the 2013-2016 EBOV outbreak in West Africa. This could help to inform the treatment of patients in future outbreaks. Additionally, this permitted the evaluation of the use of Illumina RNA-sequencing as a method of identifying those microorganisms that are potential co-infections. The detection of *P. falciparum*

presented an opportunity to do this as we have information from both from the sequenced blood and the RDT performed upon patient admission.

Data from this chapter has been published as part of the following publication:

Carroll MW, *et al*. 2017. Deep sequencing of RNA from blood and oral swab samples reveals the presence of nucleic acid from a number of pathogens in patients with acute Ebola virus disease and is consistent with bacterial translocation across the gut. mSphere 2:e00325-17. https://doi.org/10.1128/ mSphereDirect.00325-17.

The initial identification of transcripts from microorganisms was performed by colleagues at the Centre of Genomic Research, Liverpool. All downstream work is my own.

5.2 Results

5.2.1 Identification of microorganisms in the bodily fluids of patients with EVD

In order to assess the microorganisms present in the blood of acute EVD patients, RNA was extracted from the diagnostic blood sample and sequenced as previously described. The presence of EBOV was confirmed via qRT-PCR. In this study, a total of 202 individuals were included, as summarised in Table 5.1. The groups described as "Hospitalised fatalities" and "Hospitalised survivors" represent blood samples taken from patients upon arrival to the ETC suffering from acute illness. These individuals then went on to overcome the infection or succumb to it. "Community deaths" are oral swabs taken from individuals who died within the region surrounding the ETC in Guéckédou, forested Guinea, who tested positive for EBOV. Individuals who went on to survive infection were followed through their convalescence to confirm that they were virus-free. These "Convalescent controls" consist of blood samples taken from patients who once tested positive for EBOV but subsequently cleared the virus and were confirmed to be negative via qRT-PCR. These convalescent patients have been deemed to not be significantly different from a healthy Canadian cohort in terms of their transcriptomic profile (Liu *et al.*, 2017). Therefore, these convalescent patients are able to act as controls in this study. These individuals would never had had any contact with EBOV. This "never" group was recruited as part of a different study located in British Colombia (Shin *et al.*, 2014).

Table 5.3: Summary of patients involved in this study. *These have been split according to sample type and patient outcome. Here, patient age and EBOV Ct are described by the range and mean. N describes the total number of samples in each patient group.*

Patient group		**Hospitalised fatalities**	**Hospitalised survivors**	**Convalescent controls**	**Community deaths**
N		118	44	16	24
% Male		40.1	29.5	93.8	29.1
Age (years)	Range	2 mo – 80	10 mo – 68	18 – 40	1 - 70
	Mean	28.4	33.1	30.0	33.4
EBOV Ct	Range	12.1 – 26.6	15.8 – 31.6	NA	12.4 – 31.9
	Mean	17.2	21.7	NA	18.8

After sequencing, reads mapping to EBOV and the human genome were removed and Trinity employed for *de novo* assembly in order to identify transcripts mapping to potential infectious agents.

In order to evaluate this proposed technique, patient Ct value (a proxy for viral load obtained via qRT-PCR) was compared to the reads mapping to EBOV after RNA-sequencing (Figure 5.2). A linear relationship is observed between the reciprocal Ct value and the logged read values (linear regression performed, $R^2 = 0.53$, $P<0.0001$). This relationship led to the hypothesis that a greater number of reads mapping to a microorganism corresponds to a greater abundance of that microorganism in the blood.

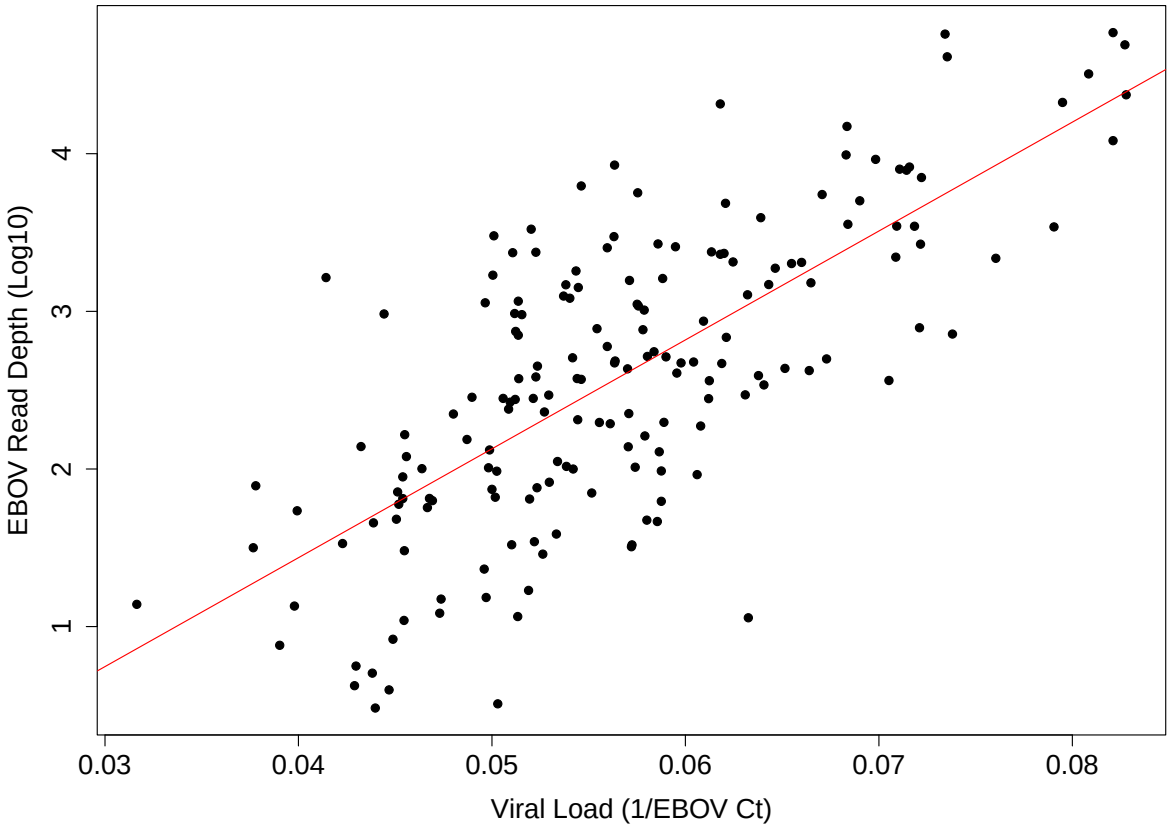


Figure 5.2: Linear relationship between viral load and sequence depth mapping to EBOV. *Viral load was determined using the reciprocal of the Ct value following qRT-PCR and sequence depth (log10 number of reads) mapping to the EBOV genome. Linear regression performed in R (red line),* $R^2 = 0.53$, $P<0.0001$.

A number of pathogens were identified as being present in these patients (a selection shown in Table 5.2, all transcripts available to download from https://doi.org/10.1128/mSphereDirect.00325-17). Reads mapping to EBOV were found in every sample, thus providing support for the effectiveness of Illumina sequencing as a platform for the identification of microorganisms present in the extracted RNA of whole blood. The identity of these microorganisms is highly supported (Table 5.2).

Table 5.4: Sequences from microorganisms identified from the blood of patients with EVD.

Microorganism	Protein	% identity	AA length	E value
EBOV	L protein	96.8	2,212	0.00E+00
GB-C	Polyprotein	99.1	328	1.50E−190
HHV-4	BALF2	100	100	1.9E−51
	LF2 protein	100	166	4.9E−91
	BALF5	98.8	172	4.4E−88
	BALF4	100	257	6.6E−142
	DNA polymerase catalytic subunit	99.2	131	2.3E−71
	BHRF1	100	121	7.40E-66
	BARF0	99.2	120	6.5E−62
	Early antigen D	100	118	2E−60
	gp110 precursor	100	95	1.3E−49
	A73 protein	100	94	6.1E−50
	LF1	98.9	92	1.1E−48
	Putative BHLF1 protein	98.7	75	4.8E−42
	BFRF1	100	71	4.1E−32
	BMRF2	100	68	7.6E−31
	dUTPase	98.5	68	3E−32
	BFRF2	100	67	4.2E−32
	LF3 protein	100	45	2E-16
	K15	100	37	5.6E-13
Human immunodeficiency virus 1	Envelope glycoprotein	94.2	69	6.30E-30

Escherichia coli	Glycosyl transferase family 2	100	583	0.00E+00
	Leucyl-tRNA synthetase	100	383	9.40E−224
Lactobacillus fermentum	DNA-directed RNA polymerase subunit beta	100	722	0.00E+00
	6-Phosphogluconate dehydrogenase	100	457	3.70E−260
Candida albicans	Acetyl coenzyme A carboxylase	100	504	2.50E−293
Streptococcus pneumoniae	Transposase DDE domain protein	100	493	2.90E−292
Streptococcus sanguinis	Collagen-binding protein	100	439	2.80E−251
Streptococcus parasanguinis	Amino acid transporter	100	402	8.80E−228
Myroides spp.	GMP synthase	100	508	3.20E−292
	Diaminopimelate decarboxylase	100	388	9.30E−223
Haemophilus influenzae	Phosphoribosylformylglycinamidine synthase	100	500	1.80E−289
Haemophilus parainfluenzae	l,d-Transpeptidase	100	393	3.80E−224
Acinetobacter baumannii	2-Isopropylmalate synthase	100	479	1.20E−276
Ureaplasma urealyticum	Cell division protein FtsH	100	478	8.20E−264
Plasmodium falciparum	Hypothetical protein C923_00328	100	511	4.50E−298
	Hypothetical protein PFFVO_04502	100	476	2.20E−270
	Hypothetical protein PFTANZ_03315	100	450	5.70E−260
	Elongation factor 1 alpha	100	433	4.60E−249
	Histidine-tRNA ligase	100	403	1.10E−235
	ATP-dependent protease HslVU,	100	410	2.10E−229

ATPase subunit

Following the identification of these microorganisms, the effect of age was investigated in several pathogens (Figure 5.3). Patients were split according to their immunological maturity (Carroll *et al.*, 2017). As expected, the peak presence of *P. falciparum* can be found in patients between 1.5 – 4 years of age, as shown previously in West Africa (Gupta *et al.*, 1994) and predicted in areas of very high transmission (Woolhouse, 1998). 100% of the individuals in this age group had reads mapping to *P. falciparum*. This remained high throughout all age groups. In previous studies, co-infection with GBV-C has been observed to have a beneficial impact on EVD patients. In our study a very small number of individuals with reads mapping to GBV-C were identified (n = 6), with all of these individuals going on to succumb to infection. The identification of HHV4 (formerly EBV), through BALF2 detection - codes for a major DNA-binding protein - signifies an active infection. Transcripts mapping to HHV4 were found in approximately 30% of individuals involved in this study, which were confirmed to be active infections due to the nature of the transcript identified. Previous studies (Sadek *et al.*, 1999; Li *et al.*, 2016) have noted that patient fatality rate was linked to age. However, our study cannot support this. The fatality rate for the over 40s was 70%, while that of 18 – 40 and the under 18s were not significantly different at 66% and 67% respectively. This lack of correlation could be due to the small number of individuals in this dataset over the age of 60, thus making conclusions difficult to draw.

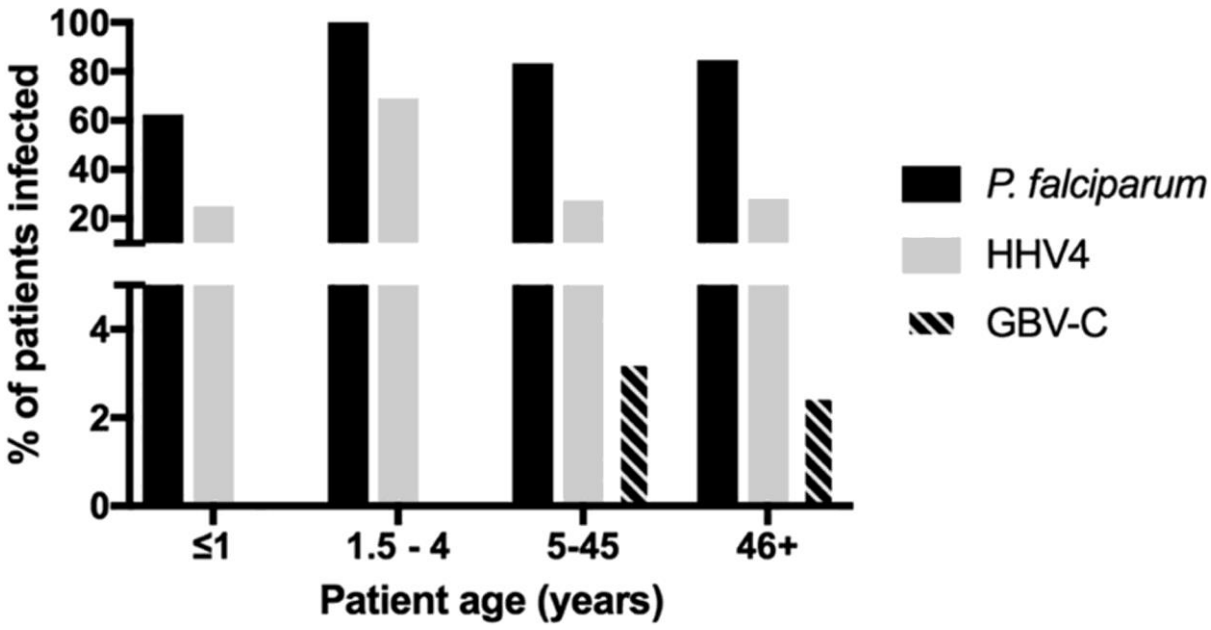


Figure 5.3: The presence of nucleic acids from certain pathogens varies with patient age. *Patients were grouped according to their immunological maturity (≤1 year [n = 8], 1.5 to 4 years [n = 13], 5 to 45 years [n = 120], and 46+ years [n = 39]), and the proportions of each group with reads mapping to P. falciparum, HHV4, and GBV-C are shown.*

After looking at the burden of disease across age ranges, the effect of potential bacterial co-infection was investigated. Here, patients were split into three groups: hospitalised fatalities, hospitalised survivors and convalescent patients. Once split, the various traces mapping to bacterial species were investigated in Figure 5.4. These traces are likely to be due to translocation of bacteria from the gut to the blood, which is indicative of gut trauma during viral infections (Balagopal *et al.*, 2008; Carroll *et al.*, 2017). Among the convalescent patients, there were very few instances of bacterial presence, suggesting that this phenomenon is associated with acute illness. In the group of patients who went on to suffer a fatal infection, there appears to be a higher proportion of individuals with reads mapping to multiple bacterial species. This is particularly remarkable when it is noted that one of the hospitalised survivors who had a high abundance of bacterial species present in their blood had lots of reads mapping to bacterial species with the ability to cause urinary tract infections – including *Pseudomonas aeruginosa*, *Prevotella amnii* and *Mycoplasma hominis* (Figure 5.4). In Chapter 3, Section 3.2 it was highlighted that a small number of individuals had reported very long symptomatic periods – up to 256 days. The potential co-infections of these individuals were of particular interest, as this was hypothesised as an explanation for this self-reported figure. However, of those individuals stating they were symptomatic for greater than 50 days, none had reads mapping to a high proportion of potential microorganisms capable of causing such morbidity.

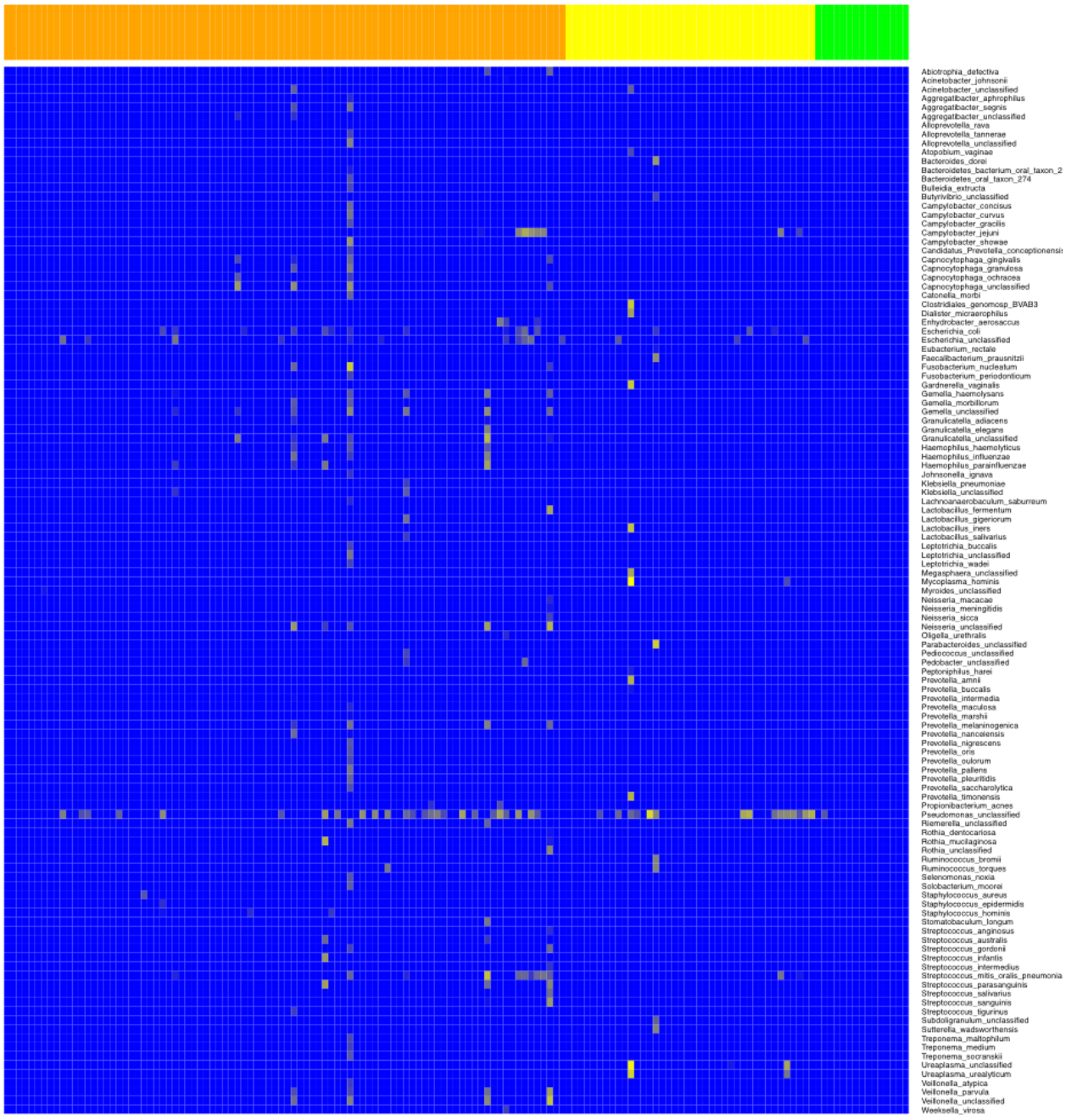


Figure 5.4: Heat map showing the bacterial species identified in the blood of acute EBOV patients (along the y-axis). *Along the x-axis is the patient group: convalescent patients in green, hospitalised survivors in yellow and hospitalised fatalities in orange (summarised in Table 5.1). Within the heatmap, the colour represents the relative read depth.*

Bacterial transcripts detected in the oral swabs of deceased EVD patients are shown in Figure 5.5. This detection was very common, with the majority of samples allowing the identification of multiple bacterial species, many of which occur naturally in the upper respiratory tract. Collaborators went on to culture five of these samples in high-containment conditions, allowing the identification of several species by MALDI-TOF (Carroll *et al.*, 2017). This cements the functionality of this pipeline in the detection of potential co-infections.

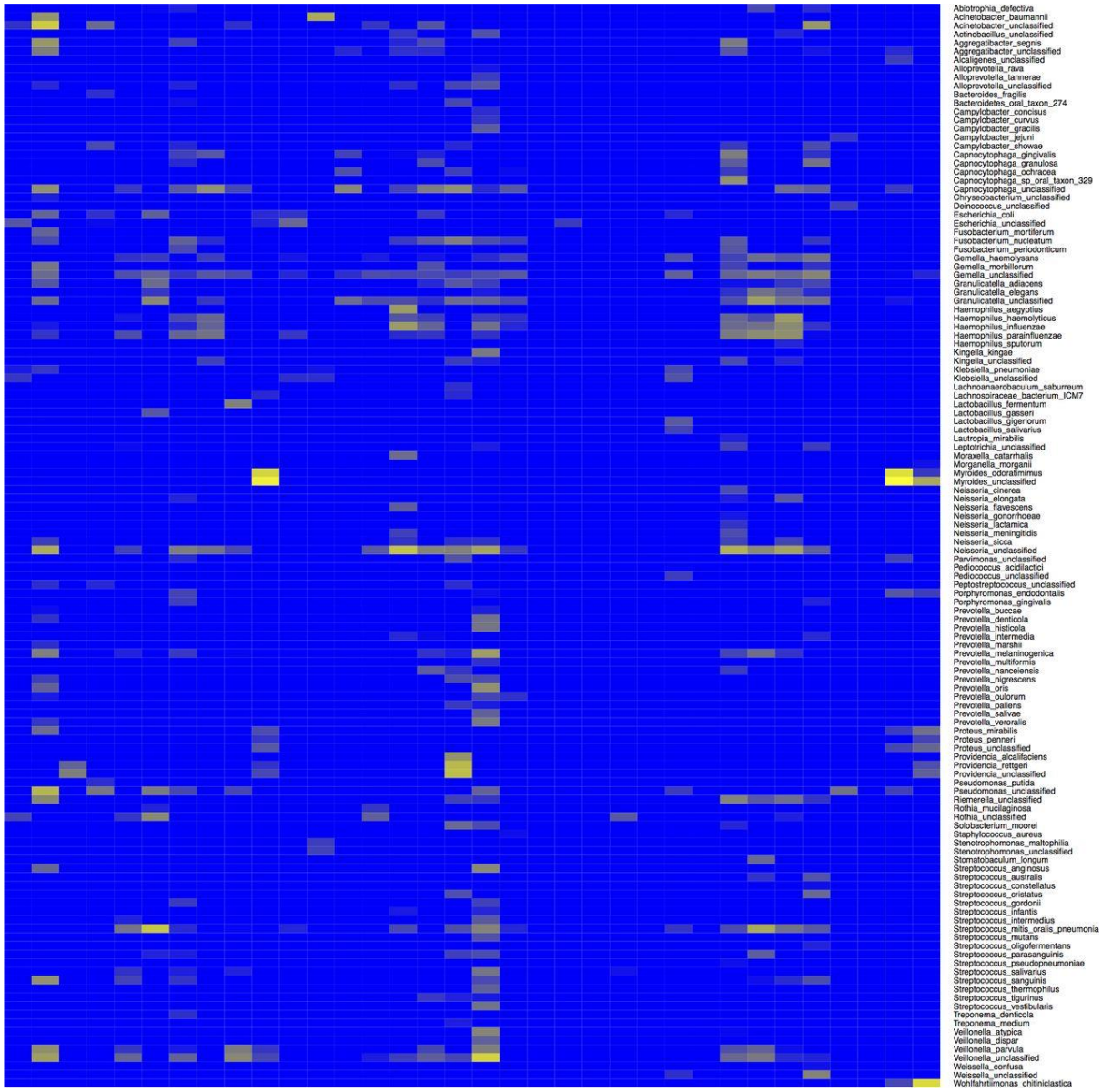


Figure 5.5: Heat map showing the transcripts identified from oral swabs of deceased individuals (n = 24). *Patient samples along the x axis with bacterial species along the y axis. The depth of colour represents the read depth, with yellow indicating a greater sequence read depth.*

5.2.2 Effect of co-infections at the level of gene expression

The differences in host response of those infected with EVD and healthy controls has been previously discussed, as well as the difference between acutely infected individuals who go on to succumb to or clear EBOV infection (Liu *et al.*, 2017). Such associated pathways include the innate immune response, blood coagulation and the acute phase response. As these have already been identified, they provide an interesting area of investigation with regards to the potential effect of the presence of bacterial transcripts in the blood. These data (Figure 5.6) suggest that the abundance of certain transcripts associated with the acute phase response could be altered according to the presence of bacterial species. The expression of ICAM-1 and TGF-β1 were significantly higher in those with no bacterial transcripts in their blood, in comparison to those who did. This trend was not observed in any of the other pathways.

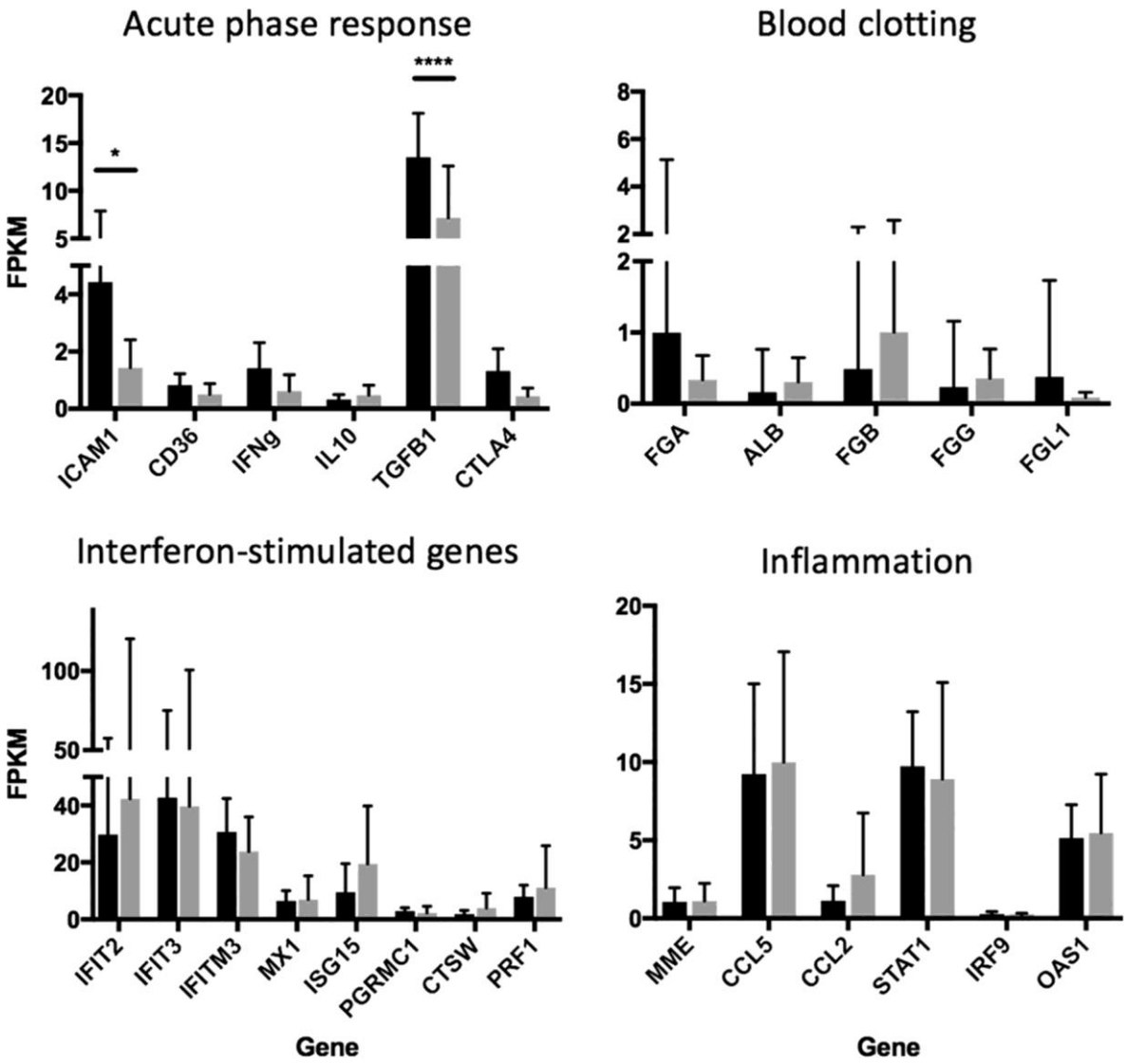


Figure 5.6: Relative abundance of transcripts mapping to host genes associated with the host acute phase response, coagulation, the interferon response and inflammation in patients with EVD in whom we could detect the transcripts from at least two bacterial species in their blood (n = 22; grey) or not (n = 9; black) (by RNA sequencing). *Relative gene expression levels are displayed as the fragments per kilobase of transcript per million mapped reads (FPKM). Graphs show mean FPKM plus standard deviation. Statistical analysis was performed using a two-way ANOVA with Sidak's multiple comparisons adjustment (Prism 7; GraphPad, CA). * p<0.05, **** p<0.0001.*

As previously mentioned, the utilisation of an RDT in the field provided an interesting comparison between diagnostic techniques. This comparison is illustrated in Figure 5.7. Here, patients have been split according to their abundance of transcripts mapping to *P. falciparum*. The patient fatality rate across these groups is shown in red, and appears to increase across these groups, with those with very high abundance of *P. falciparum* transcripts (36,000+ mapped reads) all going on to succumb to infection. Those with this high abundance likely represent patients with a severe malaria infection, probably exhibiting the many symptoms that go along with such a diagnosis. Unsurprisingly, as reads mapping to *P. falciparum* increase, the more likely that those individuals would test positive for the antigen via the RDT. There were 73 individuals who had reads mapping to *P. falciparum* and tested RDT-negative (referred to as the RDT false negative rate). 84% of individuals had reads mapping to *P. falciparum*, while only 33% tested positive via the RDT. This could potentially illustrate an under-representation of the burden of malaria in our Guinean population by the RDT. Alternatively, the RDT could be identifying severe malaria infections with greater accuracy, and the RNA-sequencing could be picking up asymptomatic or past infections.

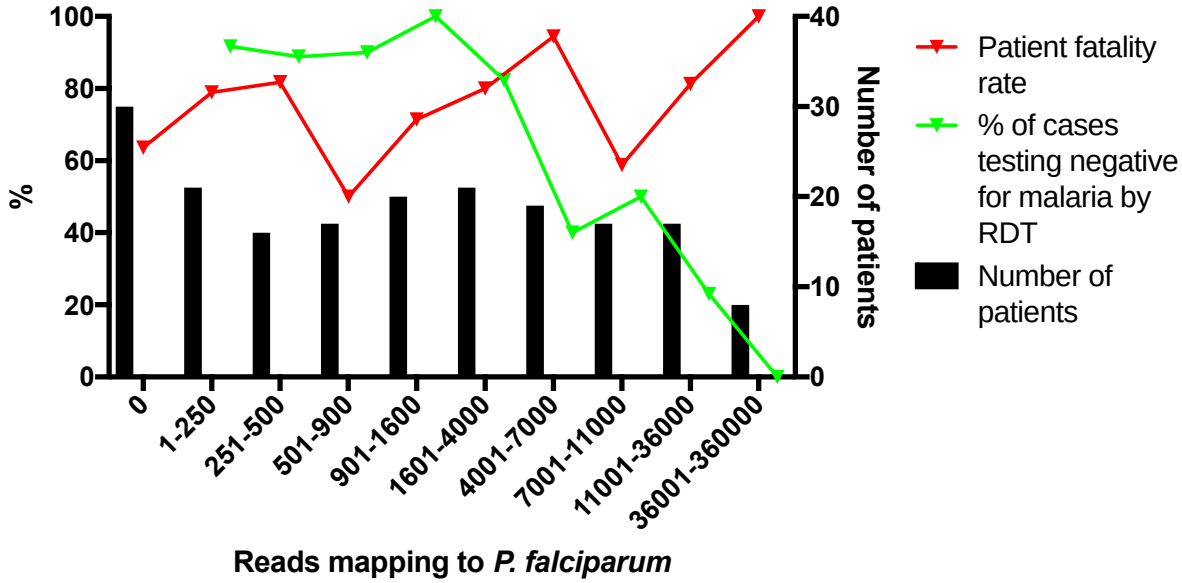


Figure 5.7: Effect of an increased number of reads mapping to P. falciparum. *The green line (associated with the left-hand y axis) shows the proportion of patients who tested negative for malaria according to the RDT but for whom reads mapping to P. falciparum were detected during RNA-seq analysis. The red line (left-hand y axis) shows the patient fatality rate. The black bars (right-hand y axis) represent the number of samples in each read group.*

In order to assess the effect of the presence of *P. falciparum* in the blood, two groups of samples were formed: (1) those testing positive for *P. falciparum* by RDT and having high number of reads mapping to the parasite, representing individuals whose immune systems have been stimulated by the co-infection and (2) those testing RDT-negative and having no reads mapping to *P. falciparum*. The latter group representing individuals who are definitely not co-infected with malaria. These groups did not differ significantly in EBOV Ct or age. This investigation included genes that have previously been implicated in being critical to the host response to severe malaria (Grangeiro de Carvalho *et al.*, 2011; Krupka *et al.*, 2012; Rommelaere *et al.*, 2015). These data suggested that there is no significant difference in the host transcript abundance between these two groups (Figure 5.8). The investigation then pivoted to consider the genes involved in blood coagulation, which have been previously highlighted as being expressed in higher levels in fatal infections (Liu *et al.*, 2017). Here, those without reads mapping to *P. falciparum* appeared to produce those transcripts in higher abundance (Figure 5.9).

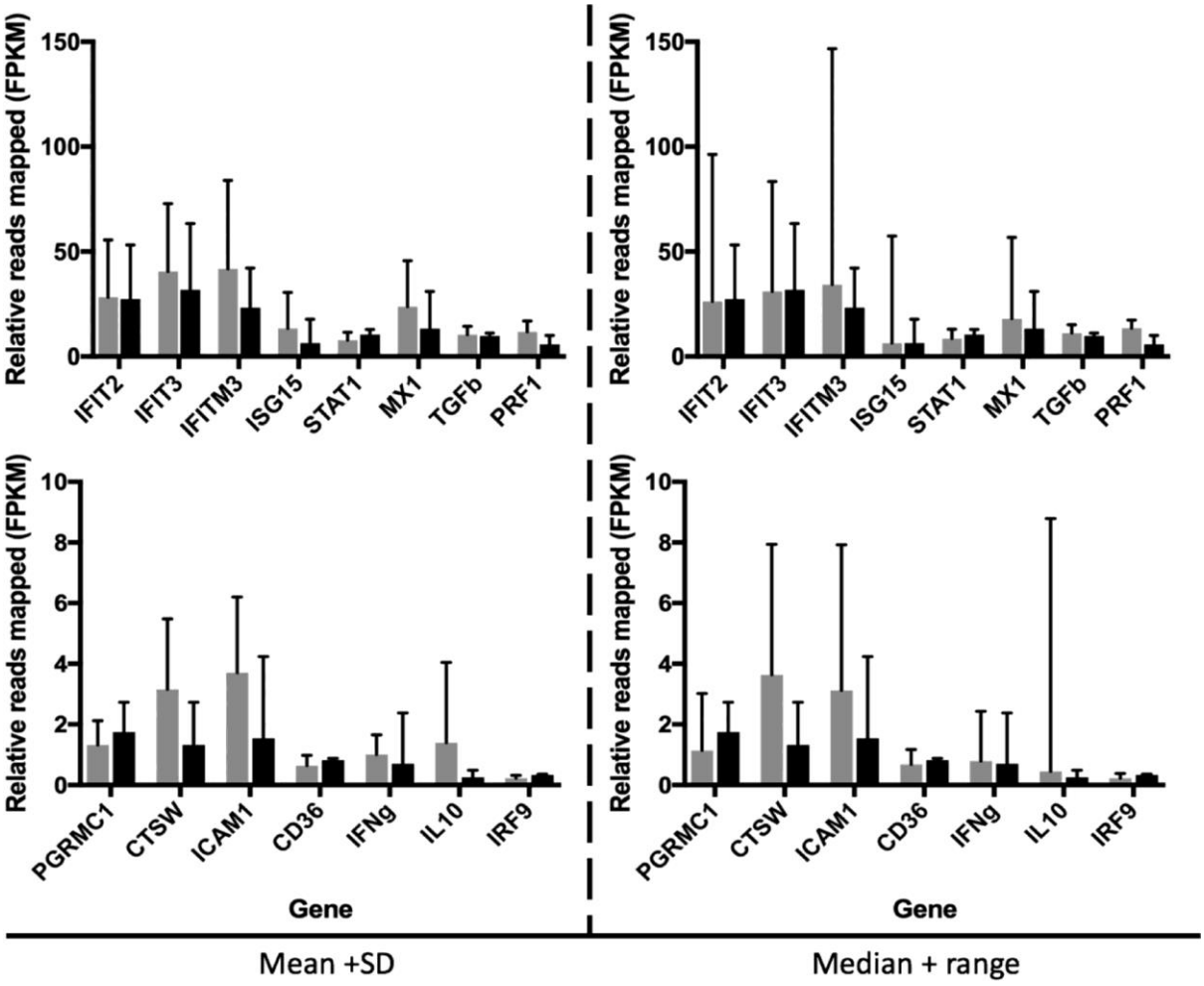


Figure 5.8: Relative abundance of transcripts mapping to host genes associated with acute malaria and the innate immune system in patients with EVD. *Comparison of two groups: those confirmed to be positive (grey; n = 10) or negative (black; n = 13) for P. falciparum (by RNA sequencing and RDT). Relative gene expression levels are displayed as the fragments per kilobase of transcript per million mapped reads (FPKM). SD = standard deviation.*

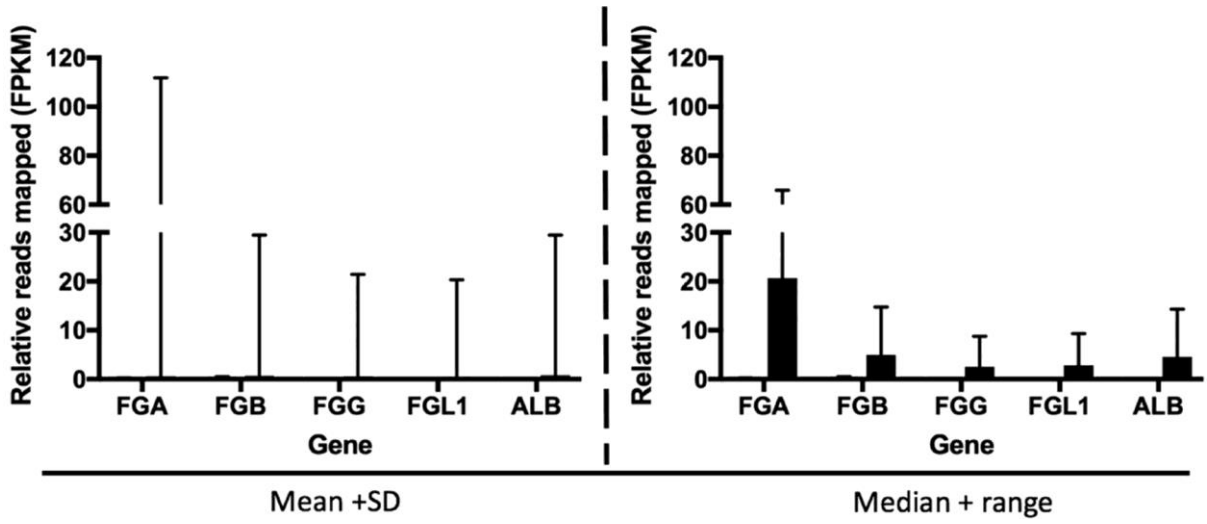


Figure 5.9: Relative abundance of transcripts mapping to host genes associated with coagulation in patients with EVD and testing positive (grey) or negative (black) for P. falciparum. *Data presented as in Figure 5.7 and the same groups used.*

In order to investigate the effect of HHV4 as a co-infection, two matched groups (as much is possible, in terms of gender, age, viral load and patient outcome) were compared – those with and without reads mapping to HHV4 (Figure 5.10).

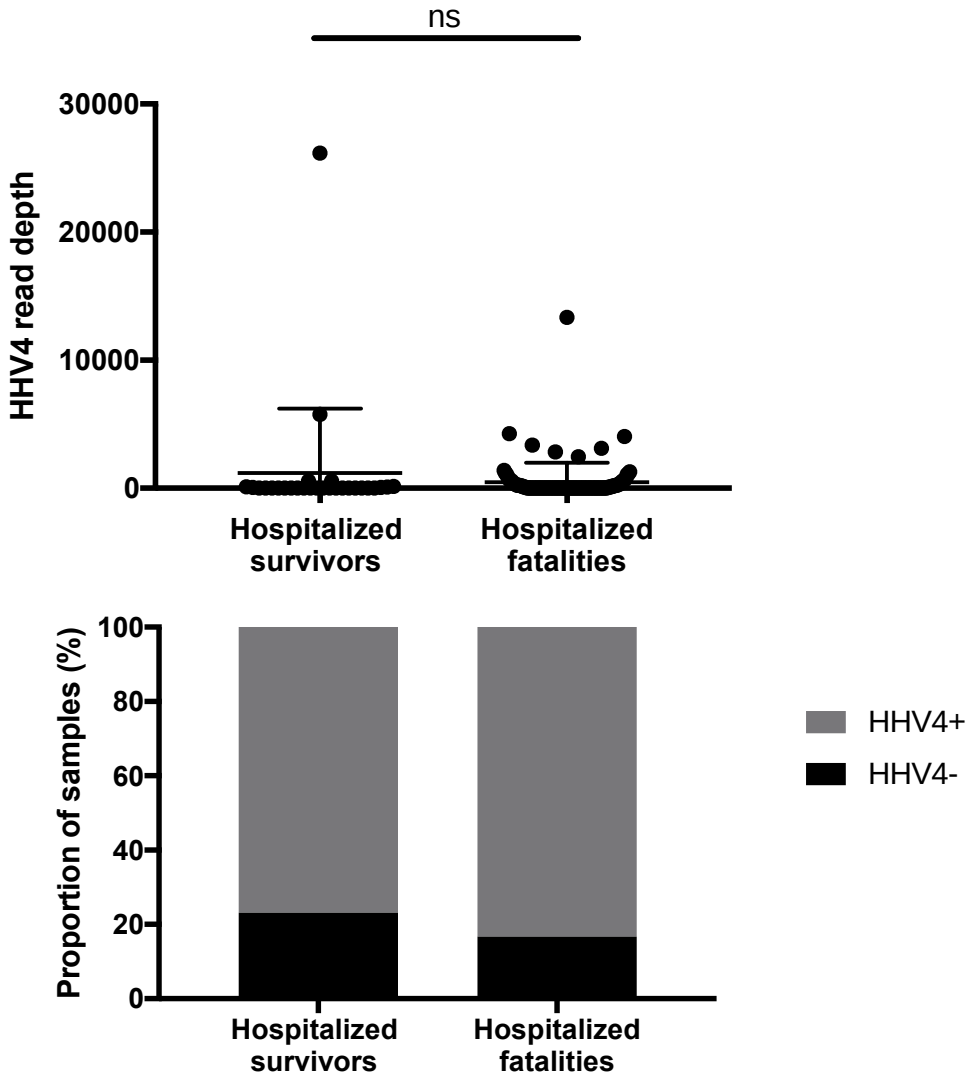


Figure 5.10: Comparing the profiles of the patients with reads mapping to HHV4 who survived or died from EVD. *There were no significant differences in terms of read depth of HHV4 hospitalised survivors and fatalities, Mann-Whitney U test performed, p > 0.2. Also, individuals with reads mapping to HHV4 did not differ in terms of fatality rate. Individuals with reads mapping to HHV4 shown in grey, without in black.*

The host transcripts were compared to that of a control group and a digital cell quantification (DCQ) performed (Figure 5.11). Upon comparison of the distinctive immune cell profiles of these groups, some differences are notable. A two-way ANOVA was performed with Sidak's multiple comparisons test, reported p-values are adjusted. Megakaryocytes ($p = 0.015$) and hematopoietic stem cells (CD38-CD34+, $p = 0.044$) were predicted to be significantly more abundant in EVD patients with HHV4 transcripts, while Naïve CD8+ T lymphocytes significantly less abundant ($p = 0.021$). Naïve B lymphocytes and neutrophils also appeared to be of lower abundance in HHV4- individuals, but this finding was not statistically significant ($p > 0.05$).

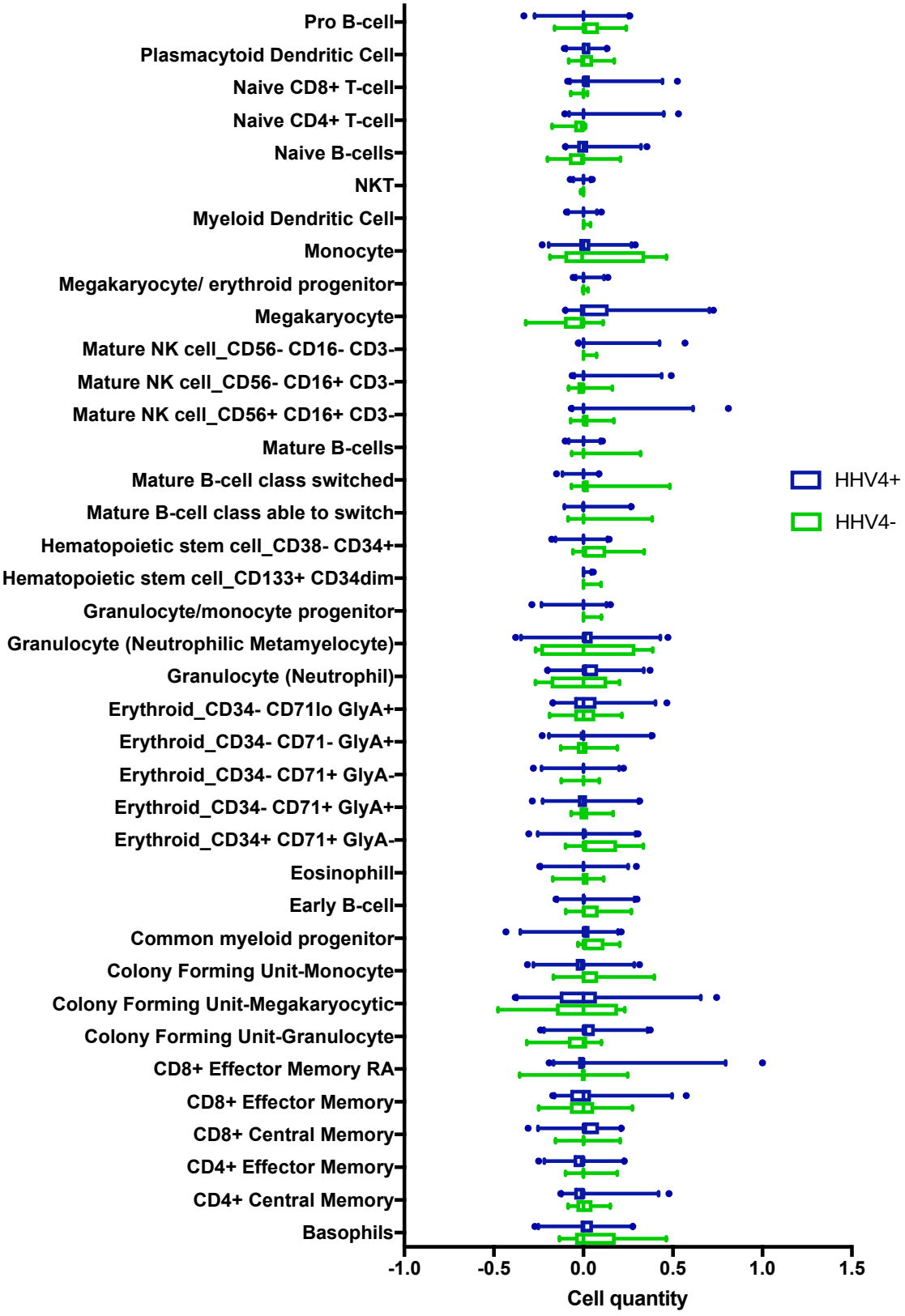


Figure 5.11: Comparison of predicted immune cell profile between HHV4 positive and negative EVD patients. *DCQ performed to compare the relative cell quantity between*

EVD-positive individuals who had reads mapping to HHV4 in their sample (HHV4+, blue; n = 14) and those who did not (HHV4-, green; n = 23). A two-way ANOVA was performed with Sidak's multiple comparisons test to compare groups using Prism 7 GraphPad (CA, USA).

5.2.3 Geographic location

As a further area for investigation, the geographic location of patients was examined (Table 5.3), with a particular focus on the forested regions of Guinea, Liberia and Sierra Leone. 25.9% of Guinea is considered to be forest area, which has decreased steadily over the past two decades from 28.8% in 1995 (The World Bank, 2016). The forested nature of Guinea has been implicated as one of the factors allowing these types outbreaks to occur due to the close contact with potential reservoir species (Weyer *et al.*, 2015). This showed that patients from the group of hospitalised fatalities had the broadest geographic spread, perhaps unsurprisingly due to this group being the largest. Geographic location had no effect on the abundance of transcripts mapping to potential pathogens. However, there do appear to be some areas in which the proportion of fatalities is higher than others, thus indicating a higher burden of disease, lacking infrastructure or extremely rural populations.

Table 5.5: Table showing geographic location of individuals involved in this study. *Where GIN – Guinea, LBR – Liberia, SLE – Sierra Leone.*

		Hospitalised fatalities	Proportion of H_F (%)	Hospitalised survivors	Proportion of H_S (%)
GIN	Macenta	27	25	7	26.9
GIN	Gueckedou	34	31.5	8	30.8
GIN	Nzerekore	5	4.63	2	7.69
GIN	Lola	2	1.85	0	0
GIN	Beyla	3	2.78	0	0
GIN	Kerouane	7	6.48	1	3.85
GIN	Faranah	1	0.93	2	7.69
GIN	Kankan	1	0.93	0	0
GIN	Siguiri	1	0.93	0	0
GIN	Kouroussa	0	0	1	3.85
GIN	Kissidougou	14	13.0	2	7.69
LBR	Foya	7	6.48	1	3.85
LBR	Voinjama	3	2.78	0	0
SLE	Kailahun	3	2.78	2	7.69

5.3 Discussion

In this chapter the presence of transcripts from microorganisms (including viral, bacterial and protozoan) in the blood of patients with EVD is investigated. This was performed using samples taken for diagnostic purposes and therefore represents an analysis that is unbiased in its sequencing approach.

The average length between the patient's first symptoms and their admission to the ETC was roughly six days (Chapter 3, Figure 3.2), suggesting that these samples represent patients who were still in the relatively early stages of infection. However, these data are based on self-reported information and therefore may not be entirely accurate - there were cases of individuals claiming to be suffering from symptoms for many months and even years. These extreme responses were not included in the analysis as these symptoms are likely to be due to other co-morbidities. The maximum incubation period for Ebola has been described as 25 days, which accounts for 99% of individuals, so this was used as a cut-off for inclusion (Eichner *et al.*, 2011). 21 days is often quoted as the incubation period for Ebola but may only account for 95% of cases. By six days, an EBOV-infected patient is likely to be in the mid-stage of infection, commonly experiencing gastrointestinal complications (Hunt *et al.*, 2015). This is categorised by diarrhoea and vomiting as well as the non-specific symptoms associated with the early stages of infection, i.e. pyrexia and lethargy (Hunt *et al.*, 2015). With these symptoms, bacterial translocation from the gut to the blood is not unlikely. Therefore, the use of blood samples in this study can provide insight into

potential co-infections. Patients were followed throughout their infection and (potential) convalescence and sequential blood samples taken. This was done to confirm that a patient was EBOV negative in their blood and therefore permitted to leave the ETC. Two consecutive blood samples testing negative for EBOV via qRT-PCR were required for patient discharge. These convalescent samples have been used as a control group in this study, providing a background profile of the transcript levels in the blood. As mentioned previously, the blood samples from acutely ill and convalescent patients were not taken in a standardised environment. During the acute stage of infection, patients can be delirious and difficult to treat, thus sample procurement becomes a more challenging task. This inequality may have impacted the microorganisms present in these samples. However, the transcriptomic profile for all acutely infected patients was not alike, so the impact of this may be negligible.

Co-infections have long been known to have the ability to influence primary infections, with the potential to significantly impact patient outcome. These relationships can be deleterious but there are cases where co-infections are deemed to be beneficial. These effects can also vary according to the geographic location, population and diagnostic method of the study. For example, higher *P. falciparum* parasitaemia has been found to be positively correlated with survival in EVD patients (Rosenke *et al.*, 2016), via the use of qRT-PCR to determine levels of the parasite. Whereas other studies have found that, especially in children between the ages of 5 and 14, a high parasitic load will regularly lead to a poor prognosis (Kerber *et al.*, 2016), upon utilisation of the RDT. This is reflected in the study described herein, where the effect of *P. falciparum* on the outcome of the patient depends on the method by which it was detected. When the

RDT was used as a determining factor, a positive result appeared to be indicative of a fatal outcome (p = 0.053; Chapter 3, Figure 3.3). However, this effect was not sustained once the RNA-sequencing data was employed to show the presence of *P. falciparum*. The explanation for this is two-fold. Firstly, a much higher proportion of individuals had reads mapping to *P. falciparum* following RNA sequencing (84%, 156 out of 186) compared to those testing positive via RDT (33%, 40 out of 121). This minimises the power behind the statistical analysis of the RNA sequencing results. The second, and perhaps the more important, explanation is that these methods differ dramatically. As the sequence depth mapping to *P. falciparum* increased, the differences between the RDT and RNA sequence data were lessened. The RDT is employed to detect patients with high levels of parasitaemia (>1000 parasites/μl blood), and so those with incidental parasitaemia/past infections may not be identified, unlike with sequencing. Following this, it is unsurprising that the detection of *P. falciparum* by RNA sequencing does not have a pronounced effect on patient outcome. Individuals with high parasitaemia are likely to have been experiencing acute malaria and suffering the accompanying symptoms. This acute phase includes a powerful proinflammatory response, which is known to be an effective dampener of viral activity. However, these responses do not always act in a synergistic fashion. Individuals with low parasitaemia are likely to have an asymptomatic infection (if there's a current infection at all) and experience little inflammation. Therefore, any interaction with a positive EBOV infection is unlikely. In order to investigate this, two groups were isolated from the sample population: (1) those testing positive for *P. falciparum* by RDT and having high number of reads mapping to the parasite and (2) those testing RDT-negative and having no reads mapping to *P. falciparum*. Comparing

these groups, there was no significant difference between the gene expression levels of genes associated with the innate immune response (Figure 5.7). This suggests that EVD is capable of completely overwhelming the immune response, so the addition of *P. falciparum* in the blood has very little effect at the level of gene transcription. There did appear to be, however, some difference in the abundance of transcripts associated with blood coagulation (Figure 5.8). This potential difference was due to a few individuals in whom the levels of gene expression were much higher. This could be due to an interruption of the host antiviral response due to the levels of high parasitaemia. Therefore, genes that are normally readily expressed in patients with an acute viral infection are dampened due to the presence of *P. falciparum* in the blood. The limitations of this study in terms of minimal samples should be noted. This is due to the finite number of samples in which no reads mapping to *P. falciparum* were found. These methods could present the opportunity to determine a cut-off point for the number of Illumina reads required to confirm the presence of a microorganism. In these data, the lowest number of reads mapping to *P. falciparum* which tested positive for the pathogen via the RDT was 198. This could provide a rough figure as to the minimum read depth at which a co-infection is considered to be present. However, this is not considering where those reads lie along the *P. falciparum* genome, which could be a critical factor.

Transcripts from a number of bacterial species were identified in blood samples of acutely ill patients with EVD (Figure 5.4), suggesting translocation of bacteria from the gut to the bloodstream. However, these were not identified in all hospitalised patients. This unbiased approach to the identification of microorganisms means that no

causative conclusions can be drawn from the data presented herein, but a microbial profile can be built. Many individuals had transcripts in their blood mapping to bacterial species capable of causing sepsis, including *Haemophilus influenzae* and *Streptococcus pneumoniae*. Furthermore, although the identities of these transcripts are well supported, the close phylogenetic relationship between bacterial species (considered to be both pathogenic and non) needs to be considered. An example of this was the identification of *P. reichenowi*. This *Plasmodium* species is known to infect chimpanzees, not humans. It is difficult to determine whether these transcripts were genuinely present in the blood samples or simply a case of mistaken identity due to the analysis pipeline. It is possible that the two species are so phylogenetically similar that they could not be differentiated. Despite these factors, the identification of numerous bacterial species associated with sepsis is concerning. This discovery supports the use of antibiotic therapy for patients with EVD. CDC guidelines suggest that antibiotics should be employed in order to distinguish nonspecific symptoms of EVD from bacterial sepsis (Uyeki *et al.*, 2016). The importance of this is clear, as sufferers of EVD are clearly at high risk of translocation of bacteria from the gut to the blood, following gastrointestinal involvement of EVD. This creates a potentially life-threatening situation. *Pseudomonas spp*. transcripts were detected in the blood of a number of patients. This Gram-negative bacterial genus contains many infectious species often found in hospital environments. In immune-compromised individuals, infection with *P. aeruginosa* can lead to sepsis and septic shock (Kielhofner *et al.*, 1992; Roy-Burman *et al.*, 2001; Kalle *et al.*, 2012; Aggarwal, 2016).

The transcripts of potential pathogens found in oral swabs of deceased individuals are unsurprisingly more abundant and diverse than those taken from blood samples (Figure 5.5). The taking of swabs post-mortem from individuals who died in the community became an essential tool in the epidemiological understanding of the outbreak, also informing burial practises. The limitations of studying these samples are plentiful, including the fact that it is impossible to tell when these bacteria would have started growing. This could have occurred during the late stages of EVD or, in fact, following the death of the individual. The presence of a number of these species were also found following mass spectrometry of the swabs, performed by colleagues (Carroll *et al.*, 2017). This helps to further validate the findings of the RNA sequencing analysis. Among the microorganisms identified by mass spectrometry were bacteria and fungi. These can be found routinely as part of the flora that constitutes the respiratory tract. In acutely ill patients, these have the potential to be opportunistic pathogens and be the cause of serious complications. These microorganisms are likely to represent a small portion of the microbial profile of the respiratory tract, as they would have to be capable of surviving harsh conditions. This includes multiple freeze-thaw cycles due to sample processing, which is known to affect the integrity of biological samples (Shabihkhani *et al.*, 2014). This illustrates one of the benefits of the use of RNA sequencing in this manner, as only extracted RNA from neutralised samples is required. From this, much information has been returned, with no need for laboratory work in high containment conditions.

In addition, there were transcripts from a number of viral species in patient samples. These were seen to vary across different age groups (Figure 5.3), but there was no

correlation between the presence of these viruses and the outcome of the patient. In previous studies, GBV-C has been seen to provide individuals with a higher chance of survival compared to their uninfected counterparts. This was not observed in this study, as all individuals with transcripts mapping to GBV-C went on to succumb to EVD. However, this group comprised of only six individuals, so no concrete conclusions can be drawn. Transcripts mapping to HHV4 were found in almost a third of patient samples, with mapping to BALF2. The downstream protein plays a crucial role in the initiation of viral entry and replication. This suggests that individuals with these transcripts were infected with HHV4. This interaction between EVD and HHV4 has been noted previously, with the reactivation of HHV4 becoming more likely as the CD4 T-cell response of the host becomes impaired (Agrati *et al.*, 2016). This effect could not be corroborated using our data (Figure 5.9). HHV4 as a coinfection with other viruses has been shown to have a deleterious effect on patient morbidity, including cytomegalovirus (CMV) (Wang *et al.*, 2010) and HIV (Chakraborty *et al.*, 2010). In our samples, however, the fatality rate was not seen to differ with the identification of HHV4 transcripts (Figure 5.10).

Human immunodeficiency virus (HIV) is a significant co-infection due to its ability to dampen the host immune response and therefore leave individuals vulnerable to other infections. In this study, only one transcript mapping to HIV-1 was detected (Table 5.2). This corresponds to the envelope glycoprotein of the RNA virus. The prevalence of HIV in Guinea is around 1% (based on 120 000 people living with HIV) (UNAIDS, 2016). At this rate, it is possible that we simply did not have more than one HIV-positive individual in our sample population. Alternatively, it is possible that our sequencing

pipeline is not appropriate for the identification of HIV, which is notoriously difficult in the absence of preamplification (Gall *et al.*, 2012).

These data provide an interesting insight into the different microorganisms and potential pathogens present in the blood of individuals suffering from EVD. The identification of bacteria associated with sepsis gives support for the treatment of patients with antibiotics upon their admission to the ETC in future outbreak situations. The detection of *P. falciparum* in individuals who tested negative via the RDT could impact how we perceive the effect of this parasite.

Chapter 6: Thesis discussion

6.1 Discussions and future perspectives

The aims of this thesis were to further the understanding of EBOV pathogenesis and potential determinants of patient outcome. This was achieved by the use of transcriptomic analysis of samples for EVD patients. Broadly, these investigations focused on the effects of the virus itself, the host response and co-infections. In order to elucidate these factors, diagnostic samples taken from the 2013-2016 EVD outbreak in West Africa were sequenced using both Illumina and Nanopore techniques as well as infected model organisms and cell lines.

This approach allowed for the comparison of acutely ill individuals at the time of sampling. Some of these individuals went on to die from infection, whilst others survived, thus creating two distinct groups: hospitalised fatalities and hospitalised survivors. Convalescent patients, testing negative for EBOV in their blood by qRT-PCR, were an invaluable control group. Upon RNAseq by Illumina, comparing the transcriptomes of these groups was facilitated. This highlighted a number of pathways that appeared to be upregulated in hospitalised patients compared to their convalescent counterparts. Many of these were associated with the acute immune response, which was to be expected due to the nature of the contrast – i.e. the comparison between any acutely ill individuals to comparatively “healthy” ones. Furthermore, these same pathways were enriched in the hospitalised fatalities compared to survivors (Figure 4.1), supporting previous findings that associated

uncontrolled inflammation and cell migration with immunopathology (Nfon *et al*., 2013). This was supported when compared to experimentally infected guinea pigs, where a general trend of increasing abundance of transcripts mapping to cytokines was observed as pathogenicity and fatality rate increased (Figure 4.12). This is critically important as it further supports the use of guinea pigs as an appropriate model organism when it comes to studying the host response to EVD. Guinea pigs are often employed for use in the first stage of testing treatments and/or vaccines prior to the involvement of NHPs. The response of NHPs to EBOV challenge has been shown to mirror the human response effectively, but the use of these models is accompanied by both ethical and financial considerations. The identification of pathways associated with a poor patient prognosis has potential clinical implications in the form of informing patient care. For example, the provision of anti-inflammatories could have a beneficial outcome if given at an appropriate time during the disease course. Of course, the context of most EVD outbreaks must be noted. These often occur in developing countries with limited resources and infrastructure and therefore implementing this knowledge in real scenarios may not be possible. Therefore, these guidelines may only be of assistance in the treatment of expatriated patients.

These enriched pathways were predicted from the identification of differentially expressed genes between the groups. This dataset allowed a linear discriminant analysis to be performed and correlates of patient outcome to be determined. With these data, patient outcome could be predicted with very high accuracy, especially in the presence of EBOV Ct (Figure 4.9). These concepts could be applied to other viral

infections and potentially more broadly to bacterial infections, with particular use in settings when pathogen burden is not a good indicator of outcome. Additionally, this LDA was constructed exclusively using samples from the West African outbreak. The ongoing outbreak in the DRC provides a potential cohort on which to test this tool. Host response to infection varies with geographical area, and so the applicability of this analysis on separate populations is still to be confirmed. The optimal way to test this would be to receive blinded RNA samples from the blood of infected individuals in the DRC that are confirmed to be EBOV-positive by qRT-PCR. Optimised targeted qRT-PCR experiments could then be performed for VCAM, TUBG1 and ISG15 and a prediction made for the outcome of the patient. If successful, a multiplex PCR could be designed and constructed to act as a predictive tool for patients with EVD. This would not delay the diagnostic period as it would consist of a single qRT-PCR run. However, the financial viability of this would have to be thoroughly investigated.

The opportunity to collect and analyse these samples provided an exciting prospect to utilise different platforms for their analysis. One such avenue was using Nanopore sequencing via the MinION. This technique was implemented during the 2013-2016 outbreak, and functioned well in the phylogenetic analysis of samples, aiding in contact tracking and the understanding of the spread of EVD (Carroll *et al.*, 2015; Quick *et al.*, 2016). For example, molecular phylogenetics allowed the beginning of the outbreak to be predicted with great accuracy, supporting the findings of a previous epidemiological study (Baize *et al.*, 2014). In this study, the aim of using this technique was to generate long reads and therefore draw further conclusions about

EBOV that could not be achieved using the data generated by an Illumina platform. This included investigating the mutations along the EBOV genome associated with patient prognosis as well understanding the true formation of DIs. However, the sub-optimal quality of the RNA extracts that were sequenced posed a significant challenge for the MinION. Due to this, the average read length generated was not much longer than those following Illumina sequencing, and coverage of the virus was poor. However, the longest reads were substantially longer than the 150 base pairs generated by Illumina platforms, thus giving hope to further investigations with this tool. Further troubleshooting and refining of this protocol could potentially produce much more successful and substantial results. The PCR kit used for the Nanopore sequencing in this study would have introduced certain biases, including a skewed preference towards shorter transcripts. In order to maximise this technology a more targeted approach would be beneficial. This could either involve EBOV-specific primers or a metagenomic approach (Kafetzopoulou *et al.*, 2018). Additionally, advances in the accuracy and ability of this platform are constantly being made, providing the opportunity to sequence RNA directly (Garalde *et al*., 2018). Initially, the error rate of this platform was prohibitively high. However, this has been substantially improved since the tool's inception and continues to develop as a viable transcriptomic device (Jain *et al.*, 2016; M. Jain *et al.*, 2018).

Although the use of the Nanopore platform did not provide as much information as initially hoped, Illumina sequencing allowed investigation into the role of the virus itself in EVD. The difference in EBOV genome variation between patient groups was significant, with surviving individuals having higher viral diversity (Figure 3.6). This is

potentially due to the host response itself, with the viral population needing to overcome population bottlenecks imposed by the immune response (Domingo et al., 2012). Certain mutations, which were identified as being common in the 2013-2016 outbreak, were found to be more common in hospitalised fatalities when compared to survivors (Figure 3.8). This could indicate that these mutations convey a fitness advantage to the virus itself, i.e. a virus capable of replicating or spreading more quickly. This is demonstrated by Dowall *et al.* (2014) in that variations in EBOV nucleotide sequences were associated with increased pathogenicity as the virus adapted to the infected guinea pigs.

One particular avenue of EBOV biology that became elucidated following the 2013-2016 outbreak was that virus persistence and sequalae. A hypothesised explanation for this is the generation of truncated viruses or DIs, which are potentially capable of interrupting viral replication. The identification of these genomes has most often been observed in in vitro infections, as was readily shown in infected THP-1 cells via the identification of fusion events (Figure 3.12). However, when these fusion events were searched for in clinical samples, very few were identified (Table 3.3). Of those identified, even fewer would be likely to be able to confer much of an evolutionary advantage due to their small size. These were expected to be found in the semen of convalescent patients, as these are the body compartments in which the virus is persisting. Therefore, DIs are potentially not the mechanism responsible for allowing EBOV to remain in the body. However, these semen samples had a lower read depth across the EBOV genome when compared to the infected THP-1 cells. This could directly impact the ability of fusion events to be detected and potential DIs to be

identified. Additionally, further investigation into different types of DIs could be beneficial, as this study only explored those formed via deletion events.

In the past, EVD was referred to as Ebola Haemorrhagic Fever. This change in nomenclature has come about due to the fact that few victims of recent outbreaks have experienced haemorrhagic symptoms (Fletcher *et al.*, 2014). One notable aspect of the 2013-2016 outbreak was the progression of patients from non-specific symptoms with fever to the gastrointestinal stage of disease (Fletcher *et al.*, 2014). This stage involves the dysfunction of the liver and kidneys. In this study, the extent of the gastrointestinal involvement is supported by the detection of transcripts from microorganisms in patient samples that would not normally be found in the blood. This suggested gut translocation of bacterial species was common in acute sufferers of the disease, while mostly absent in convalescent patients (Figure 5.4). Understanding the pathogenicity of EVD is essential not only in recognising outbreaks in their early stages, but also the handling of clinical cases. These data have important implications for the treatment of EVD patients, particularly considering antibiotic stewardship. Using this bioinformatics approach allowed EBOV to be detected in each sample. This, along with the confirmation of certain bacterial species by mass spectrometry (Carroll *et al.*, 2017), gives support for the use of bioinformatics analysis in microorganism identification.

6.1 Conclusions

An enormous amount of information has been generated through the sequencing of EBOV-positive RNA extracts. As such, only a small proportion of this study into EBOV has been performed under CL4 conditions. These data could have significant clinical implications on the guidelines of the treatment of EVD patients, for example in the provision of anti-inflammatories and/or antibiotics. One factor these analyses highlight is the benefit of personalised medicine, and the understanding that patient factors such as age can have substantial impact on their immune response and likely co-infections. However, outbreaks of EVD occur in populations that share borders with the suspected animal reservoirs of the virus – namely those in forested regions. These are usually developing countries with limited infrastructure and a dearth of education concerning infectious disease among the local population. Therefore, the opportunity to treat individuals differently in outbreak situations becomes almost impossible. A perhaps more practical solution would be to house those patients with similar requirements in close proximity. This would be beneficial for infected individuals and healthcare workers alike, especially in outbreak situations.

This thesis highlights the fact that animal studies, cell culture and the analysis of clinical samples should be considered in tandem in order to effectively study all aspects of EBOV biology.

6.2 Relevant publications with supplementary information

Liu X, Speranza E, Muñoz-Fontela C, Haldenby S, Rickett NY *et al.* 2017. Transcriptomic signatures differentiate survival from fatal outcomes in humans infected with Ebola virus. Genome Biology 18:4

Carroll MW, Haldenby S, Rickett NY *et al*. 2017. Deep sequencing of RNA from blood and oral swab samples reveals the presence of nucleic acid from a number of pathogens in patients with acute Ebola virus disease and is consistent with bacterial translocation across the gut. mSphere 2:e00325-17. https://doi.org/10.1128/mSphereDirect.00325-17.

Works cited

Aggarwal, M. (2016) 'A Rare Case of Fatal Endocarditis and Sepsis Caused by Pseudomonas aeruginosa in a Patient with Chronic Renal Failure', *Journal of Clinical and Diagnostic Research*, 10(7), pp. 12–13. doi: 10.7860/JCDR/2016/20220.8175.

Agrati, C. *et al.* (2016) 'Longitudinal characterization of dysfunctional T cell-activation during human acute Ebola infection.', *Cell death & disease*. Nature Publishing Group, 7(3), p. e2164. doi: 10.1038/cddis.2016.55.

Akerlund, E., Prescott, J. B. and Tampellini, L. (2015) 'Shedding of Ebola Virus in an Asymptomatic Pregnant Woman.', *The New England journal of medicine*, 372(25), pp. 2467–9. doi: 10.1056/NEJMc1503275.

Aljabr, W. *et al.* (2016) 'Investigating the Influence of Ribavirin on Human Respiratory Syncytial Virus RNA Synthesis by Using a High-Resolution', *Journal of Virology*, 90(10), pp. 4876–4888. doi: 10.1128/JVI.02349-15.Editor.

Altboum, Z. *et al.* (2014) 'Digital cell quantification identifies global immune cell dynamics during influenza infection', *Mol Syst Biol.*, 10. doi: 10.1002/msb.134947.

Andrew, M. *et al.* (1992) 'Maturation of the hemostatic system during childhood.', *Blood*, 80(8), pp. 1998–2005. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1391957.

Auer, H. *et al.* (2014) 'The effects of frozen tissue storage conditions on the integrity of RNA and protein', *Biotechnic and Histochemistry*. Taylor & Francis, 89(7), pp. 518–528. doi: 10.3109/10520295.2014.904927.

Auffray, C., Sieweke, M. H. and Geissmann, F. (2009) 'Blood monocytes: development, heterogeneity, and relationship with dendritic cells', *Annu Rev Immunol.*, 27. doi:

10.1146/annurev.immunol.021908.132557.

Baize, S. *et al.* (1999) 'Defective humoral responses and extensive intravascular apoptosis are associated with fatal outcome in Ebola virus-infected patients.', *Nature medicine*, 5(4), pp. 423–426. doi: 10.1038/7422.

Baize, S. *et al.* (2014) 'Emergence of Zaire Ebola Virus Disease in Guinea - Preliminary Report.', *The New England journal of medicine*, 371(15), pp. 1418–1425. doi: 10.1056/NEJMoa1404505.

Balagopal, A. *et al.* (2008) 'Human Immunodeficiency Virus-Related Microbial Translocation and Progression of Hepatitis C', *Gastroenterology*, 135(1), pp. 226–233. doi: 10.1053/j.gastro.2008.03.022.

Basler, C. F. *et al.* (2000) 'The Ebola virus VP35 protein functions as a type I IFN antagonist', *Proceedings of the National Academy of Sciences*, 97(22), pp. 12289–12294. doi: 10.1073/pnas.220398297.

Basler, C. F. and Amarasinghe, G. K. (2009) 'Evasion of Interferon Responses by Ebola and Marburg Viruses', *Journal of Interferon & Cytokine Research*, 29(9), pp. 511–520. doi: 10.1089/jir.2009.0076.

Bausch, D. G. *et al.* (2007) 'Assessment of the Risk of Ebola Virus Transmission from Bodily Fluids and Fomites', *Journal of Infectious Diseases*, 2699(Suppl 2), pp. 2–7. doi: 10.1086/520545.

Bosio, C. M. *et al.* (2003) 'Ebola and Marburg Viruses Replicate in Monocyte-Derived Dendritic Cells without Inducing the Production of Cytokines and Full Maturation', *The Journal of Infectious Diseases*, 188(11), pp. 1630–1638. doi: 10.1086/379199.

Boss, I. W. and Renne, R. (2010) 'Viral miRNAs: Tools for immune evasion', *Current Opinion in Microbiology*. Elsevier Ltd, 13(4), pp. 540–545. doi: 10.1016/j.mib.2010.05.017.

Bosworth, A. *et al.* (2017) 'A comparison of host gene expression signatures associated with infection in vitro by the Makona and Ecran (Mayinga) variants of Ebola virus', *Nature Publishing Group*. Nature Publishing Group, (July 2016), pp. 1–15. doi: 10.1038/srep43144.

Breman, J. G. *et al.* (1980) 'A Search for Ebola Virus in Animals in the Democratic Republic of the Congo and Cameroon : Ecologic , Virologic , and Serologic Surveys , 1979 – 1980', pp. 1979–1980.

Buchfink, B., Xie, C. and Huson, D. H. (2015) 'Fast and sensitive protein alignment using DIAMOND', *Nature Methods*, 12(1), pp. 59–60. doi: 10.1038/nmeth.3176.

Caballero, I. S. *et al.* (2016) 'In vivo Ebola virus infection leads to a strong innate response in circulating immune cells', *BMC Genomics*. BMC Genomics, 17(1), pp. 1–13. doi: 10.1186/s12864-016-3060-0.

Calain, P., Monroe, M. C. and Nichol, S. T. (1999) 'Ebola virus defective interfering particles and persistent infection', *Virology*, 262(1), pp. 114–128. Available at: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed4&NEWS=N&AN=1999337759.

Cane, P. A., Matthews, D. A. and Pringle, C. R. (1994) 'Analysis of respiratory syncytial virus strain variation in successive epidemics in one city.', *Journal of Clinical Microbiology*, 32(1), pp. 1 LP – 4. Available at: http://jcm.asm.org/content/32/1/1.abstract.

Capone, F. *et al.* (2014) 'Cytokinome profile evaluation in patients with hepatitis C virus infection', *World Journal of Gastroenterology*, 20(28), pp. 9261–9269. doi: 10.3748/wjg.v20.i28.9261.

Carl, S. *et al.* (2002) 'Modulation of Different Human Immunodeficiency Virus Type 1 Nef Functions during Progression to AIDS', *Journal of Virology*, 75(8), pp. 3657–3665.

doi: 10.1128/jvi.75.8.3657-3665.2001.

Carroll, M. W. *et al.* (2015) 'Temporal and spatial analysis of the 2014-2015 Ebola virus outbreak in West Africa', *Nature*, 524(7563). doi: 10.1038/nature14594.

Carroll, M. W. *et al.* (2017) 'Deep Sequencing of RNA from Blood and Oral Swab Samples Reveals the Presence of Nucleic Acid from a Number of Pathogens in Patients with Acute Ebola Virus Disease and Is Consistent with Bacterial Translocation across the Gut', *mSphere*, 2(4), pp. 1–19. doi: 10.1128/mSphereDirect.00325-17.

CDC (2019a) *40 Years of Ebola Virus Disease around the World*. Available at: https://www.cdc.gov/vhf/ebola/history/chronology.html (Accessed: 1 August 2019).

CDC (2019b) *Number of Cases and Deaths in Guinea, Liberia, and Sierra Leone during the 2014-2016 West Africa Ebola Outbreak*. Available at: https://www.cdc.gov/vhf/ebola/history/2014-2016-outbreak/case-counts.html (Accessed: 7 May 2019).

Chakraborty, N. *et al.* (2010) 'Incidence of multiple Herpesvirus infection in HIV seropositive patients, a big concern for Eastern Indian scenario', *Virology Journal*, 7(1), p. 147. doi: 10.1186/1743-422X-7-147.

Cherif, M. S. *et al.* (2017) 'Ebola virus disease in children during the 2014-2015 epidemic in Guinea: a nationwide cohort study.', *European journal of pediatrics*. Germany, 176(6), pp. 791–796. doi: 10.1007/s00431-017-2914-z.

Chertow, D. S. *et al.* (2014) 'Ebola Virus Disease in West Africa — Clinical Manifestations and Management', *New England Journal of Medicine*, 371(22), pp. 2054–2057. doi: 10.1056/NEJMp1411794.

Christie, A. *et al.* (2015) 'Possible Sexual Transmission of Ebola Virus — Liberia , 2015', *MMWR*, 64(17), pp. 479–481. doi: mm6417a2 [pii].

Cimini, E. *et al.* (2017) 'Different features of Vδ2 T and NK cells in fatal and non-fatal

human Ebola infections', *PLoS Neglected Tropical Diseases*, 11(5), pp. 1–11. doi: 10.1371/journal.pntd.0005645.

Clark, D. V *et al.* (2012) 'Long-term sequelae after Ebola virus disease in Bundibugyo , Uganda : a retrospective cohort study', *The Lancet Infectious Diseases*. Elsevier Ltd, 15(8), pp. 905–912. doi: 10.1016/S1473-3099(15)70152-0.

Connolly, B. M. *et al.* (1999) 'Pathogenesis of Experimental Ebola Virus Infection in Guinea Pigs', *The Journal of Infectious Diseases*, 179(s1), pp. S203–S217. doi: 10.1086/514305.

Conrady, C. D. *et al.* (2013) 'IFN-α-driven CCL2 production recruits inflammatory monocytes to infection site in mice', *Mucosal Immunology*. Nature Publishing Group, 6(1), pp. 45–55. doi: 10.1038/mi.2012.46.

Cooper, M. A. *et al.* (2001) 'Human natural killer cells : a unique innate immunoregulatory role for the CD56 bright subset Human natural killer cells : a unique innate immunoregulatory role for the CD56 bright subset', *Immunobiology*, 97(10), pp. 3146–3151. doi: 10.1182/blood.V97.10.3146.

Crotty, S., Cameron, C. E. and Andino, R. (2001) 'RNA virus error catastrophe: Direct molecular test by using ribavirin', *Proceedings of the National Academy of Sciences*, 98(12), pp. 6895–6900. doi: 10.1073/pnas.111085598.

Culley, F. J. *et al.* (2006) 'Differential Chemokine Expression following Respiratory Virus Infection Reflects Th1- or Th2-Biased Immunopathology', *Journal of Virology*, 80(9), pp. 4521–4527. doi: 10.1128/JVI.80.9.4521.

Darby, S. C. *et al.* (1996) 'Importance of age at infection with HIV-1 for survival and development of AIDS in UK haemophilia population', *Lancet*, 347(9015), pp. 1573–1579. doi: 10.1016/S0140-6736(96)91073-9.

Deen, G. F. *et al.* (2015) 'Ebola RNA persistence in semen of Ebola Virus Disease

survivors - preliminary report', *N Engl J Med*.

Delisle, B. *et al.* (2012) 'Porcine reproductive and respiratory syndrome virus diversity of Eastern Canada swine herds in a large sequence dataset reveals two hypervariable regions under positive selection', *Infection, Genetics and Evolution*. Elsevier B.V., 12(5), pp. 1111–1119. doi: 10.1016/j.meegid.2012.03.015.

Diallo, B. *et al.* (2016) 'Resurgence of Ebola Virus Disease in Guinea Linked to a Survivor with Virus Persistence in Seminal Fluid for More Than 500 Days', *Clinical Infectious Diseases*, 63(10), pp. 1353–1356. doi: 10.1093/cid/ciw601.

Dimmock, N. J. *et al.* (2008) 'Influenza Virus Protecting RNA: an Effective Prophylactic and Therapeutic Antiviral', *Journal of Virology*, 82(17), pp. 8570–8578. doi: 10.1128/jvi.00743-08.

Dinçer, E. *et al.* (2017) 'Generic amplification and next generation sequencing reveal Crimean-Congo hemorrhagic fever virus AP92-like strain and distinct tick phleboviruses in Anatolia, Turkey', *Parasites and Vectors*, 10(1), pp. 1–16. doi: 10.1186/s13071-017-2279-1.

Domingo, E. and Holland, J. J. (1997) 'RNA VIRUS MUTATIONS', *Annual Review of Microbiology*, 51, pp. 151–178.

Domingo, E., Sheldon, J. and Perales, C. (2012) 'Viral Quasispecies Evolution', *Microbiology and Molecular Biology Reviews*, 76(2), pp. 159–216. doi: 10.1128/MMBR.05023-11.

Dowall, S. D. *et al.* (2014) 'Elucidating variations in the nucleotide sequence of Ebola virus associated with increasing pathogenicity', *Genome Biology*, 15(11), pp. 1–12. doi: 10.1186/s13059-014-0540-x.

Ebihara, H. *et al.* (2011) 'Host response dynamics following lethal infection of rhesus macaques with Zaire ebolavirus', *J Infect Dis.*, 204(SUPPL. 3), pp. 14–16. doi:

10.1093/infdis/jir336.

Eichner, M., Dowell, S. F. and Firese, N. (2011) 'Incubation Period of Ebola Hemorrhagic Virus Subtype Zaire', *Osong Public Health and Research Perspectives*. Elsevier, 2(1), pp. 3–7. doi: 10.1016/j.phrp.2011.04.001.

Emond, R. T. D. *et al.* (1977) 'A Case Of Ebola Virus Infection', *The British Medical Journal*. BMJ, 2(6086), pp. 541–544. Available at: http://www.jstor.org/stable/20415676.

Falasca, L. *et al.* (2015) 'Molecular mechanisms of Ebola virus pathogenesis: focus on cell death.', *Cell death and differentiation*. Nature Publishing Group, 22(8), pp. 1250–1259. doi: 10.1038/cdd.2015.67.

Farci, P. *et al.* (2000) 'The Outcome of Acute Hepatitis C Predicted by the Evolution of the Viral Quasispecies', *Science*, 288(5464), pp. 339 LP – 344. doi: 10.1126/science.288.5464.339.

Farci, P. *et al.* (2006) 'Evolution of hepatitis C viral quasispecies and hepatic injury in perinatally infected children followed prospectively.', *Proceedings of the National Academy of Sciences of the United States of America*. United States, 103(22), pp. 8475–8480. doi: 10.1073/pnas.0602546103.

Feldmann, H. and Geisbert, T. W. (2011) 'Ebola haemorrhagic fever', *The Lancet*. Elsevier Ltd, 377(9768), pp. 849–862. doi: 10.1016/S0140-6736(10)60667-8.

Fensterl, V. and Sen, G. C. (2009) 'Interferons and viral infections', *BioFactors*, 35(1), pp. 14–20. doi: 10.1002/biof.6.

Fischer, N. *et al.* (2015) 'Evaluation of Unbiased Next-Generation Sequencing of RNA (RNA-seq) as a Diagnostic Method in Influenza Virus-Positive Respiratory Samples.', *Journal of Clinical Microbiology*, 53(7), pp. 2238–50. doi: 10.1128/JCM.02495-14.

Fitzpatrick, G. *et al.* (2015) 'The Contribution of Ebola Viral Load at Admission and

Other Patient Characteristics to Mortality in a Médecins Sans Frontières Ebola Case Management Centre, Kailahun, Sierra Leone, June-October 2014', *Journal of Infectious Diseases*, 212(10), pp. 1752–1758. doi: 10.1093/infdis/jiv304.

Fletcher, T. E., Fowler, R. A. and Beeching, N. J. (2014) 'Understanding organ dysfunction in Ebola virus disease', *Intensive Care Medicine*, 40(12), pp. 1936–1939. doi: 10.1007/s00134-014-3515-1.

Formenty, P., Boesch, C., *et al.* (1999) 'Ebola virus outbreak among wild chimpanzees living in a rain forest of Cote d'Ivoire.', *The Journal of infectious diseases*. United States, 179 Suppl, pp. S120-6. doi: 10.1086/514296.

Formenty, P., Hatz, C., *et al.* (1999) 'Human infection due to Ebola virus, subtype Cote d'Ivoire: clinical and biologic presentation.', *The Journal of infectious diseases*. United States, 179 Suppl, pp. S48-53. doi: 10.1086/514285.

Formenty, P. *et al.* (2003) '[Outbreak of Ebola hemorrhagic fever in the Republic of the Congo, 2003: a new strategy?].', *Medecine tropicale : revue du Corps de sante colonial*. France, 63(3), pp. 291–295.

Fowler, R. A. *et al.* (2014) 'CRITICAL CARE PERSPECTIVE Caring for Critically Ill Patients with Ebola Virus Disease Perspectives from West Africa', 190, pp. 733–737. doi: 10.1164/rccm.201408-1514CP.

Frishberg, A. *et al.* (2016) 'ImmQuant: a user-friendly tool for inferring immune cell-type composition from gene-expression data', *Bioinformatics*, 32(24), pp. 3842–3843. doi: 10.1093/bioinformatics/btw535.

Fülöp, T., Larbi, A. and Pawelec, G. (2013) 'Human T cell aging and the impact of persistent viral infections', *Frontiers in Immunology*, 4(SEP), pp. 1–9. doi: 10.3389/fimmu.2013.00271.

Gall, A. *et al.* (2012) 'Universal amplification, next-generation sequencing, and

assembly of HIV-1 genomes', *Journal of Clinical Microbiology*, 50(12), pp. 3838–3844. doi: 10.1128/JCM.01516-12.

Garalde, D. R. *et al.* (2018) 'Highly parallel direct RNA sequencing on an array of nanopores', *Nature Methods*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 15, p. 201. Available at: https://doi.org/10.1038/nmeth.4577.

Geisbert, T. W. *et al.* (2000) 'Apoptosis induced in vitro and in vivo during infection by Ebola and Marburg viruses.', *Laboratory investigation; a journal of technical methods and pathology*. United States, 80(2), pp. 171–186.

Geisbert, T. W., Young, H. A., *et al.* (2003) 'Mechanisms Underlying Coagulation Abnormalities in Ebola Hemorrhagic Fever: Overexpression of Tissue Factor in Primate Monocytes/Macrophages Is a Key Event', *The Journal of Infectious Diseases*, 188(11), pp. 1618–1629. doi: 10.1086/379724.

Geisbert, T. W., Hensley, L. E., *et al.* (2003) 'Pathogenesis of Ebola Hemorrhagic Fever in Cynomolgus Macaques', *The American journal of pathology*. American Society for Investigative Pathology, 163(6), pp. 2347–2370. doi: 10.1016/S0002-9440(10)63591-2.

Genin, M. *et al.* (2015) 'M1 and M2 macrophages derived from THP-1 cells differentially modulate the response of cancer cells to etoposide', *BMC Cancer*, 15(1), p. 577. doi: 10.1186/s12885-015-1546-9.

Gire, S. K. *et al.* (2014) 'Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak', *Science*, 345(6202), pp. 1369–1372.

Glynn, J. R. *et al.* (2017) 'Asymptomatic infection and unrecognised Ebola virus disease in Ebola-affected households in Sierra Leone: a cross-sectional study using a new non-invasive assay for antibodies to Ebola virus', *The Lancet Infectious Diseases*. doi: 10.1016/S1473-3099(17)30111-1.

Goldstein, T. *et al.* (2018) 'The discovery of Bombali virus adds further support for bats as hosts of ebolaviruses', *Nature Microbiology*, 3(10), pp. 1084–1089. doi: 10.1038/s41564-018-0227-2.

Gomez, C. R. *et al.* (2008) 'Innate immunity and aging', *Experimental Gerontology*, 43, pp. 718–728. doi: 10.1016/j.exger.2008.05.016.

González Saldaña, N. *et al.* (2012) 'Clinical and laboratory characteristics of infectious mononucleosis by Epstein-Barr virus in Mexican children', *BMC Research Notes*, 5(1), p. 361. doi: 10.1186/1756-0500-5-361.

Graci, J. D. and Cameron, C. E. (2006) 'Mechanisms of action of ribavirin against distinct viruses', *Reviews in Medical Virology*, 16(1), pp. 37–48. doi: 10.1002/rmv.483.

Grangeiro de Carvalho, E. *et al.* (2011) 'Plasmodium falciparum-Infected Erythrocytes and IL-12/IL-18 Induce Diverse Transcriptomes in Human NK Cells: IFN-α/β Pathway versus TREM Signaling', *PLOS ONE*. Public Library of Science, 6(9), p. e24963. Available at: https://doi.org/10.1371/journal.pone.0024963.

Guan, M. *et al.* (2012) 'Three different functional microdomains in the hepatitis C virus hypervariable region 1 (HVR1) mediate entry and immune evasion', *Journal of Biological Chemistry*, 287(42), pp. 35631–35645. doi: 10.1074/jbc.M112.382341.

Gupta, S. *et al.* (1994) 'Parasite virulence and disease patterns in Plasmodium falciparum malaria.', *Proceedings of the National Academy of Sciences*, 91(9), pp. 3715–3719. doi: 10.1073/pnas.91.9.3715.

Guzman, M. G. *et al.* (2002) 'Original Report Effect of age on outcome dengue 2 infections of secondary', *Int J Infect Dis*, 6(2), pp. 118–124.

Haas, B. J. *et al.* (2013) 'De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis.', *Nature protocols*, 8(8), pp. 1494–512. doi: 10.1038/nprot.2013.084.

Haniffa, M. *et al.* (2012) 'Human tissues contain CD141hi cross-presenting dendritic cells with functional homology to mouse CD103+ nonlymphoid dendritic cells', *Immunity.*, 37. doi: 10.1016/j.immuni.2012.04.012.

Hayman, D. T. S. *et al.* (2010) 'Long-Term Survival of an Urban Fruit Bat Seropositive for Ebola and Lagos Bat Viruses', *PLoS ONE*, 5(8), pp. 2008–2010. doi: 10.1371/journal.pone.0011978.

Hensley, L. E. *et al.* (2012) 'Proinflammatory response during Ebola virus infection of primate models: possible involvement of the tumor necrosis factor receptor superfamily', *Immunology Letters 80*, 80, pp. 169–179. Available at: http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=530&showFR=1.

Hofmann-Winkler, H. *et al.* (2015) 'Comparative Analysis of Host Cell Entry of Ebola Virus from Sierra Leone, 2014, and Zaire, 1976', *Journal of Infectious Diseases*, 212(Suppl 2), pp. S172–S180. doi: 10.1093/infdis/jiv101.

Hunt, L. *et al.* (2015) 'Clinical presentation, biochemical, and haematological parameters and their association with outcome in patients with Ebola virus disease: An observational cohort study', *The Lancet Infectious Diseases*. Elsevier Ltd, 15(11), pp. 1292–1299. doi: 10.1016/S1473-3099(15)00144-9.

Hutchinson, K. L. and Rollin, P. E. (2007) 'Cytokine and chemokine expression in humans infected with Sudan Ebola virus.', *The Journal of infectious diseases*, 196 Suppl, pp. S357–S363. doi: 10.1086/520611.

Jain, C. *et al.* (2018) 'A Fast Approximate Algorithm for Mapping Long Reads to Large Reference Databases', *Journal of Computational Biology*. Mary Ann Liebert, Inc., publishers, 25(7), pp. 766–779. doi: 10.1089/cmb.2018.0036.

Jain, M. *et al.* (2016) 'The Oxford Nanopore MinION: Delivery of nanopore sequencing

to the genomics community', *Genome Biology*. Genome Biology, 17(1), pp. 1–11. doi: 10.1186/s13059-016-1103-0l.

Jain, M. *et al.* (2018) 'Linear assembly of a human centromere on the Y chromosome', *Nature Biotechnology*. The Author(s), 36, p. 321. Available at: https://doi.org/10.1038/nbt.4109.

Johnson, E. *et al.* (1995) 'Lethal experimental infections of rhesus monkeys by aerosolized Ebola virus', *Int. J. Exp. Path.*, 76, pp. 227–236.

Jong, M. D. De *et al.* (2006) 'Fatal outcome of human influenza A (H5N1) is associated with high viral load and hypercytokinemia', *Nature medicine*, 12(10), pp. 1203–1207. doi: 10.1038/nm1477.Fatal.

Kafetzopoulou, L. E. *et al.* (2018) 'Assessment of Metagenomic MinION and Illumina sequencing as an approach for the recovery of whole genome sequences of chikungunya and dengue viruses directly from clinical samples.', *Euro Surveill.*, 23(50), pp. 1–13. doi: 10.1101/355560.

Kalle, M. *et al.* (2012) 'Host Defense Peptides of Thrombin Modulate Inflammation and Coagulation in Endotoxin-Mediated Shock and Pseudomonas aeruginosa Sepsis', *PLoS ONE*, 7(12). doi: 10.1371/journal.pone.0051313.

Kanehisa, M. and Goto, S. (2000) 'KEGG: kyoto encyclopedia of genes and genomes.', *Nucleic acids research*. England, 28(1), pp. 27–30. doi: 10.1093/nar/28.1.27.

Kerber, R. *et al.* (2016) 'Analysis of Diagnostic Findings from the European Mobile Laboratory in Guéckédou, Guinea, March 2014 Through March 2015', *Journal of Infectious Diseases*, 214(March 2014), pp. S250–S257. doi: 10.1093/infdis/jiw269.

Khrustalev, V. V., Barkovsky, E. V. and Khrustaleva, T. A. (2015) 'Local Mutational Pressures in Genomes of Zaire Ebolavirus and Marburg Virus', *Advances in Bioinformatics*, 2015, pp. 1–14. doi: 10.1155/2015/678587.

Kielhofner, M. *et al.* (1992) 'Life-threatening Pseudomonas aeruginosa infections in patients with human immunodeficiency virus infection', *Clin Infect Dis*, 14(2), pp. 403–411.

Kim, D. *et al.* (2013) 'TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions', *Genome Biol.*, 14. doi: 10.1186/gb-2013-14-4-r36.

Kim, D. and Salzberg, S. L. (2011) 'TopHat-Fusion: an algorithm for discovery of novel fusion transcripts', *Genome Biology*, 12(8), p. R72. doi: 10.1186/gb-2011-12-8-r72.

Kreuels, B. *et al.* (2014) 'A Case of Severe Ebola Virus Infection Complicated by Gram-Negative Septicaemia', *The New England Journal of Medicine*, 371(25), pp. 1–8. doi: 10.1056/NEJMoa1411677.

Krupka, M. *et al.* (2012) 'Mild Plasmodium falciparum Malaria following an Episode of Severe', *Infection and Immunity*, 80(3), pp. 1150–1155. doi: 10.1128/IAI.06008-11.

Kuhn, J. H. *et al.* (2014) 'Nomenclature- and database-compatible names for the Two Ebola virus variants that emerged in guinea and the Democratic Republic of the Congo in 2014', *Viruses*, 6(11), pp. 4760–4799. doi: 10.3390/v6114760.

Lace, M. J. *et al.* (2008) 'The E8 E2 Gene Product of Human Papillomavirus Type 16 Represses Early Transcription and Replication but Is Dispensable for Viral Plasmid Persistence in Keratinocytes', *Journal of Virology*, 82(21), pp. 10841–10853. doi: 10.1128/jvi.01481-08.

Lahaye, X. *et al.* (2009) 'Functional Characterization of Negri Bodies (NBs) in Rabies Virus-Infected Cells: Evidence that NBs Are Sites of Viral Transcription and Replication', *Journal of Virology*, 83(16), pp. 7948–7958. doi: 10.1128/JVI.00554-09.

Lamb, L. *et al.* (2015) 'Bacterial co-infection is rare in patients with Ebola virus disease in a military Ebola virus disease treatment unit in Sierra Leone', *Journal of Infection*.

Elsevier Ltd, 71(3), pp. 406–407. doi: 10.1016/j.jinf.2015.03.004.

Lamontagne, F. *et al.* (2018) 'Evidence-based guidelines for supportive care of patients with Ebola virus disease', *Lancet*, 391(10121), pp. 700–708. doi: 10.1016/S0140-6736(17)31795-6.Evidence-based.

Langmead, B. and Salzberg, S. L. (2012) 'Fast gapped-read alignment with Bowtie 2', *Nat Methods.*, 9. doi: 10.1038/nmeth.1923.

Lanini, S. *et al.* (2018) 'Relationship between Viremia and Specific Organ Damage in Ebola Patients: A Cohort Study', *Clinical Infectious Diseases*, 66(1), pp. 36–44. doi: 10.1093/cid/cix704.

Lauck, M. *et al.* (2014) 'GB virus C co-infections in West African Ebola patients.', *Journal of virology*, 89(December), p. JVI.02752-14-. doi: 10.1128/JVI.02752-14.

Leroy, E. M. *et al.* (2001) 'Early immune responses accompanying human asymptomatic Ebola infections', *Clinical and Experimental Immunology*, 124, pp. 453–460. doi: 10.1046/j.1365-2249.2001.01517.x.

Leroy, E. M. *et al.* (2004) 'Multiple Ebola Virus Transmission Events and Rapid Decline of Central African Wildlife', *Science*, 303(5656), pp. 387 LP – 390. doi: 10.1126/science.1092528.

Leroy, E. M. *et al.* (2005) 'Fruit bats as reservoirs of Ebola virus.', *Nature*, 438(December), pp. 575–576. doi: 10.1038/438575a.

Li, D. *et al.* (2011) 'Defective interfering viral particles in acute dengue infections', *PLoS ONE*, 6(4). doi: 10.1371/journal.pone.0019447.

Li, J. *et al.* (2016) 'Age and Ebola viral load correlate with mortality and survival time in 288 Ebola virus disease patients', *International Journal of Infectious Diseases*. International Society for Infectious Diseases, 42, pp. 34–39. doi: 10.1016/j.ijid.2015.10.021.

Liu, X. *et al.* (2017) 'Transcriptomic signatures differentiate survival from fatal outcomes in humans infected with Ebola virus', *Genome Biology*. Genome Biology, 18(1), p. 4. doi: 10.1186/s13059-016-1137-3.

Lofts, L. L. *et al.* (2011) 'Key Genomic Changes Necessary for an In Vivo Lethal Mouse Marburgvirus Variant Selection Process', *Journal of Virology*, 85(8), pp. 3905–3917. doi: 10.1128/JVI.02372-10.

Lupfer, C. *et al.* (2013) 'Receptor interacting protein kinase 2-mediated mitophagy regulates inflammasome activation during virus infection', *Nature Immunology*, 14(5), pp. 480–488. doi: 10.1038/ni.2563.

Lyon, G. M. *et al.* (2014) 'Clinical care of two patients with Ebola virus disease in the United States', *N Engl J Med.*, 371. doi: 10.1056/NEJMoa1409838.

Mahalanabis, D. *et al.* (1973) 'Oral fluid therapy of cholera among Bangladesh refugees', *Johns Hopkins Medical Journal*, 132(4), pp. 197–205.

Mahanty, S. *et al.* (2003) 'Cutting Edge: Impairment of Dendritic Cells and Adaptive Immunity by Ebola and Lassa Viruses', *The Journal of Immunology*, 170(6), pp. 2797–2801. doi: 10.4049/jimmunol.170.6.2797.

Malakhova, O. A. and Zhang, D. E. (2008) 'ISG15 inhibits Nedd4 ubiquitin E3 activity and enhances the innate antiviral response', *Journal of Biological Chemistry*, 283(14), pp. 8783–8787. doi: 10.1074/jbc.C800030200.

Mari Saez, A. *et al.* (2015) 'Investigating the zoonotic origin of the West African Ebola epidemic', *EMBO Molecular Medicine*, 7(1), pp. 17–23. doi: 10.15252/emmm.201404792.

Marriott, A. C. and Dimmock, N. J. (2010) 'Defective interfering viruses and their potential as antiviral agents', *Reviews in medical virology*, 20, pp. 51–62. doi: 10.1002/rmv.

Martin, J. N. *et al.* (1995) 'Effect of older age on survival in human immunodeficiency virus (HIV) disease', *American Journal of Epidemiology*, 142(11), pp. 1221–1230. doi: 10.1093/oxfordjournals.aje.a117581.

Martínez, I. and Melero, J. A. (2002) 'A model for the generation of multiple A to G transitions in the human respiratory syncytial virus genome: Predicted RNA secondary structures as substrates for adenosine deaminases that act on RNA', *Journal of General Virology*, 83(6), pp. 1445–1455. doi: 10.1099/0022-1317-83-6-1445.

Martinez, O. *et al.* (2013) 'A mutation in the Ebola virus envelope glycoprotein restricts viral entry in a host species- and cell-type-specific manner.', *Journal of virology*. United States, 87(6), pp. 3324–3334. doi: 10.1128/JVI.01598-12.

Martini, G. A. and Schmidt, H. A. (1968) 'Spermatogene Übertragung des „Virus Marburg``', *Klinische Wochenschrift*, 46(7), pp. 398–400. doi: 10.1007/BF01734141.

Marzi, A. *et al.* (2015) 'Delayed disease progression in cynomolgus macaques infected with Ebola virus Makona strain', *Emerg Infect Dis.*, 21. doi: 10.3201/eid2110.150259.

Marzi, A. *et al.* (2018) 'Recently Identified Mutations in the Ebola Virus-Makona Genome Do Not Alter Pathogenicity in Animal Models', *Cell Reports*, 23(6), pp. 1–8. doi: 10.4172/2157-7633.1000305.Improved.

McElroy, A. K. *et al.* (2014a) 'Biomarker correlates of survival in pediatric patients with Ebola virus disease', *Emerg Infect Dis.*, 20. doi: 10.3201/eid2010.140430.

McElroy, A. K. *et al.* (2014b) 'Ebola hemorrhagic Fever: novel biomarker correlates of clinical outcome', *J Infect Dis.*, 210, pp. 558–566. doi: 10.1093/infdis/jiu088.

McElroy, A. K. *et al.* (2015) 'Human Ebola virus infection results in substantial immune activation', *Proc Natl Acad Sci U S A.*, 112(15), pp. 4719–4724. doi: 10.1073/pnas.1502619112.

McElroy, A. K. *et al.* (2016) 'Kinetic analysis of biomarkers in a cohort of US patients

with Ebola virus disease', *Clin Infect Dis.*, 63. doi: 10.1093/cid/ciw334.

McWilliam Leitch, E. C. and McLauchlan, J. (2013) 'Determining the Cellular Diversity of Hepatitis C Virus Quasispecies by Single-Cell Viral Sequencing', *Journal of Virology*, 87(23), pp. 12648 LP – 12655. doi: 10.1128/JVI.01602-13.

Morchang, A. *et al.* (2011) 'Cell death gene expression profile: Role of RIPK2 in dengue virus-mediated apoptosis', *Virus Research*. Elsevier B.V., 156, pp. 25–34. doi: 10.1016/j.virusres.2010.12.012.

Nakayama, E. and Saijo, M. (2013) 'Animal models for Ebola and Marburg virus infections', *Frontiers in Microbiology*, 4(267), pp. 1–20. doi: 10.3389/fmicb.2013.00267.

Nanbo, A. *et al.* (2013) 'The spatio-temporal distribution dynamics of Ebola virus proteins and RNA in infected cells', *Scientific reports*, 3(1206). doi: 10.1038/srep01206.

Nfon, C. K. *et al.* (2013) 'Immunopathogenesis of Severe Acute Respiratory Disease in Zaire ebolavirus-Infected Pigs', *PLoS ONE*, 8(4). doi: 10.1371/journal.pone.0061904.

Nkengasong, J. N. and Onyebujoh, P. (2018) 'Response to the Ebola virus disease outbreak in the Democratic Republic of the Congo', *The Lancet*, 391(10138), pp. 2395–2398. doi: 10.1016/S0140-6736(18)31326-6.

Noppornpanth, S. *et al.* (2007) 'Characterization of Hepatitis C Virus Deletion Mutants Circulating in Chronically Infected Patients', *Journal of Virology*, 81(22), pp. 12496–12503. doi: 10.1128/JVI.01059-07.

O'Shea, M. K. *et al.* (2016) 'Case report: A health care worker with Ebola virus disease and adverse prognostic factors treated in Sierra Leone', *American Journal of Tropical Medicine and Hygiene*, 94(4), pp. 829–832. doi: 10.4269/ajtmh.15-0461.

Okumura, A., Pitha, P. M. and Harty, R. N. (2008) 'ISG15 inhibits Ebola VP40 VLP budding in an L-domain-dependent manner by blocking Nedd4 ligase activity',

Proceedings of the National Academy of Sciences, 105(10), pp. 3974–3979. doi: 10.1073/pnas.0710629105.

Ostler, T., Davidson, W. and Ehl, S. (2002) 'Virus clearance and immunopathology by CD8+ T cells during infection with respiratory syncytial virus are mediated by IFN-γ', *European Journal of Immunology*, 32(8), pp. 2117–2123. doi: 10.1002/1521-4141(200208)32:8<2117::AID-IMMU2117>3.0.CO;2-C.

Padilla-Carlin, D. J., McMurray, D. N. and Hickey, A. J. (2008) 'The guinea pig as a model of infectious diseases', *Comparative Medicine*, 58(4), pp. 324–340.

Park, D. J. *et al.* (2015) 'Ebola Virus Epidemiology, Transmission, and Evolution during Seven Months in Sierra Leone', *Cell*, 161(7), pp. 1516–1526. doi: 10.1016/j.cell.2015.06.007.

Pawelec, G. *et al.* (2012) 'The impact of CMV infection on survival in older humans', *Current Opinion in Immunology*. Elsevier Ltd, 24(4), pp. 507–511. doi: 10.1016/j.coi.2012.04.002.

Del Portillo, A. *et al.* (2011) 'Multiploid Inheritance of HIV-1 during Cell-to-Cell Infection', *Journal of Virology*, 85(14), pp. 7169 LP – 7176. doi: 10.1128/JVI.00231-11.

Pourrut, X. *et al.* (2009) 'Large serological survey showing cocirculation of Ebola and Marburg viruses in Gabonese bat populations , and a high seroprevalence of both viruses in Rousettus aegyptiacus', *BioMed Central*, 10, pp. 1–10. doi: 10.1186/1471-2334-9-159.

QIAGEN (2016) *GeneGlobe Data Analysis Center*. Available at: https://www.qiagen.com/gb/shop/genes-and-pathways/data-analysis-center-overview-page/ (Accessed: 21 June 2017).

Qin, Z. (2012) 'The use of THP-1 cells as a model for mimicking the function and regulation of monocytes and macrophages in the vasculature', *Atherosclerosis*, 221(1),

pp. 2–11. doi: https://doi.org/10.1016/j.atherosclerosis.2011.09.003.

Qiu, X. *et al.* (2014) 'Establishment and Characterization of a Lethal Mouse Model for the Angola Strain of Marburg Virus', *Journal of Virology*, 88(21), pp. 12703–12714. doi: 10.1128/JVI.01643-14.

Quick, J. *et al.* (2016) 'Real-time, portable genome sequencing for Ebola surveillance', *Nature*, 530(7589). doi: 10.1038/nature16996.

Ramesh, A. *et al.* (2018) 'Etiology of fever in Ugandan children: identification of microbial pathogens using metagenomic next-generation sequencing and IDseq, a platform for unbiased metagenomic analysis', 44, pp. 1–42. doi: 10.1113/expphysiol.2014.081265.This.

Randall, R. E. and Goodbourn, S. (2008) 'Interferons and viruses: An interplay between induction, signalling, antiviral responses and virus countermeasures', *Journal of General Virology*, 89(1), pp. 1–47. doi: 10.1099/vir.0.83391-0.

Reed, D. S. *et al.* (2004) 'Depletion of peripheral blood T lymphocytes and NK cells during the course of ebola hemorrhagic Fever in cynomolgus macaques', *Viral Immunol*, 17(3), pp. 390–400. doi: 10.1089/0882824041857058.

Reed, P. E. *et al.* (2014) 'A new approach for monitoring ebolavirus in wild great apes.', *PLoS neglected tropical diseases*. United States, 8(9), p. e3143. doi: 10.1371/journal.pntd.0003143.

Reid, S. P. *et al.* (2006) 'Ebola virus VP24 binds karyopherin alpha1 and blocks STAT1 nuclear accumulation.', *Journal of virology*, 80(11), pp. 5156–5167. doi: 10.1128/JVI.02349-05.

Rénia, L. and Potter, S. M. (2006) 'Co-infection of malaria with HIV: An immunological perspective', *Parasite Immunology*, 28(11), pp. 589–595. doi: 10.1111/j.1365-3024.2006.00903.x.

Rieger, T. *et al.* (2016) 'Evaluation of RealStar Reverse Transcription – Polymerase Chain Reaction Kits for Filovirus Detection in the Laboratory and Field', *The Journal of Infectious Diseases*, 214(Suppl 3), pp. 243–249. doi: 10.1093/infdis/jiw246.

Rima, B. K. and Duprex, W. P. (2005) 'Molecular mechanisms of measles virus persistence', *Virus Research*, 111, pp. 132–147. doi: 10.1016/j.virusres.2005.04.005.

Robinson, M. D., McCarthy, D. J. and Smyth, G. K. (2010) 'edgeR: a Bioconductor package for differential expression analysis of digital gene expression data', *Bioinformatics*, 26. doi: 10.1093/bioinformatics/btp616.

Rommelaere, S. *et al.* (2015) 'Serum Pantetheinase / Vanin Levels Regulate Erythrocyte Homeostasis and Severity of Malaria', *The American Journal of Pathology*. American Society for Investigative Pathology, 185(11), pp. 3039–3052. doi: 10.1016/j.ajpath.2015.07.011.

Rosenke, K. *et al.* (2016) 'Plasmodium Parasitemia Associated With Increased Survival in Ebola Virus-Infected Patients.', *Clinical infectious diseases*, p. ciw452. doi: 10.1093/cid/ciw452.

Roy-Burman, a *et al.* (2001) 'Type III protein secretion is associated with death in lower respiratory and systemic Pseudomonas aeruginosa infections.', *The Journal of infectious diseases*, 183(12), pp. 1767–1774. doi: 10.1086/320737.

Rubins, K. H. *et al.* (2007) 'The temporal program of peripheral blood gene expression in the response of nonhuman primates to Ebola hemorrhagic fever', *Genome Biology*, 8. doi: 10.1186/gb-2007-8-8-r174.

Ruibal, P. *et al.* (2016) 'Unique human immune signature of Ebola virus disease in Guinea', *Nature*. Nature Publishing Group, 533(7601), pp. 100–104. doi: 10.1038/nature17949.

Rustagi, A. and Gale, M. (2014) 'Innate antiviral immune signaling, viral evasion and

modulation by HIV-1', *Journal of Molecular Biology*. Elsevier B.V., 426(6), pp. 1161–1177. doi: 10.1016/j.jmb.2013.12.003.

Sadek, R. F. *et al.* (1999) 'Ebola Hemorrhagic Fever, Democratic Republic of the Congo, 1995: Determinants of Survival', *The Journal of Infectious Diseases*, 179(s1), pp. S24–S27. doi: 10.1086/514311.

Safronetz, D., Geisbert, T. W. and Feldmann, H. (2013) 'Animal models for highly pathogenic emerging viruses', *Current Opinion in Virology*. Elsevier B.V., 3(2), pp. 205–209. doi: 10.1016/j.coviro.2013.01.001.

Samuel, C. E. (2011) 'Adenosine deaminases acting on RNA (ADARs) are both antiviral and proviral', *Virology*, 411(2), pp. 180–193. doi: 10.1016/j.virol.2010.12.004.

Saraya, A. W. *et al.* (2016) 'Normocellular CSF in herpes simplex encephalitis', *BMC Research Notes*. BioMed Central, 9(1), pp. 1–7. doi: 10.1186/s13104-016-1922-9.

Schieffelin, J. S. *et al.* (2014) 'Clinical illness and outcomes in patients with Ebola in Sierra Leone', *N Engl J Med.*, 371(22), pp. 2092–2100. doi: 10.1056/NEJMoa1411680.

Severson, W. E. *et al.* (2002) 'Ribavirin Causes Error Catastrophe during Hantaan Virus Replication', *Journal of Virology*, 77(1), pp. 481–488. doi: 10.1128/jvi.77.1.481-488.2003.

Shabihkhani, M. *et al.* (2014) 'The procurement, storage, and quality assurance of frozen blood and tissue biospecimens in pathology, biorepository, and biobank settings', *Clinical Biochemistry*. The Canadian Society of Clinical Chemists, 47(4–5), pp. 258–266. doi: 10.1016/j.clinbiochem.2014.01.002.

Shabman, R. S. *et al.* (2014) 'Deep Sequencing Identifies Noncanonical Editing of Ebola and Marburg Virus RNAs in Infected Cells', *mBio*, 6925(6), pp. 1–11. doi: 10.1128/mBio.02011-14.Editor.

Sheehy, A. M. *et al.* (2002) 'Isolation of a human gene that inhibits HIV-1 infection and

is suppressed by the viral Vif protein', *Nature*, 418(6898), pp. 646–650. doi: 10.1038/nature00939.

Shin, H. *et al.* (2014) 'Variation in RNA-Seq transcriptome profiles of peripheral whole blood from healthy individuals with and without globin depletion', *PLoS One.*, 9. doi: 10.1371/journal.pone.0091041.

Simon-Lorière, E. *et al.* (2015) 'Distinct lineages of Ebola virus in Guinea during the 2014 West African epidemic', *Nature*, 524, pp. 102–104. doi: 110.1038/nature14612.

Sissoko, D. *et al.* (2016) 'Experimental treatment with Favipiravir for Ebola Virus Disease (the JIKI Trial): A historically controlled, single-arm proof-of-concept trial in Guinea', *PLoS Med.*, 13(3), p. e1001967. doi: 10.1371/journal.pmed.1001967.

Sissoko, D. *et al.* (2017) 'Persistence and clearance of Ebola virus RNA from seminal fluid of Ebola virus disease survivors: a longitudinal analysis and modelling study', *The Lancet Global Health*. The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY license, 5(1), pp. e80–e88. doi: 10.1016/S2214-109X(16)30243-1.

Smith, L. M. *et al.* (2013) 'Interferon-beta therapy prolongs survival in rhesus macaque models of Ebola and Marburg hemorrhagic fever', *J Infect Dis.*, 208. doi: 10.1093/infdis/jis921.

Smither, S. J. *et al.* (2016) 'Ebola virus Makona shows reduced lethality in an immune-deficient mouse model', *J Infect Dis.*, 214. doi: 10.1093/infdis/jiw145.

Solana, R. *et al.* (2012) 'Innate immunosenescence: Effect of aging on cells and receptors of the innate immune system in humans', *Seminars in Immunology*. Elsevier Ltd, 24(5), pp. 331–341. doi: 10.1016/j.smim.2012.04.008.

Solomon, T. *et al.* (2012) 'Management of suspected viral encephalitis in adults - Association of British Neurologists and British Infection Association National

Guidelines', *Journal of Infection*. Elsevier Ltd, 64(4), pp. 347–373. doi: 10.1016/j.jinf.2011.11.014.

Stoop, J. N. *et al.* (2005) 'Regulatory T cells contribute to the impaired immune response in patients with chronic hepatitis B virus infection', *Hepatology*, 41(4), pp. 771–778. doi: 10.1002/hep.20649.

Stricher, F. *et al.* (2013) 'HSPA8/HSC70 chaperone protein: Structure, function, and chemical targeting', *Autophagy*, 9(12), pp. 1937–1954. doi: 10.4161/auto.26448.

Suspene, R. *et al.* (2011) 'Double-Stranded RNA Adenosine Deaminase ADAR-1-Induced Hypermutated Genomes among Inactivated Seasonal Influenza and Live Attenuated Measles Virus Vaccines', *Journal of Virology*, 85(5), pp. 2458–2462. doi: 10.1128/JVI.02138-10.

The.World.Bank (2016) *Guinea indicators*, *Forest area (% of land area)*. Available at: https://data.worldbank.org/indicator/AG.LND.FRST.ZS?locations=GN.

Thorburn, F. *et al.* (2015) 'The use of next generation sequencing in the diagnosis and typing of respiratory infections', *Journal of Clinical Virology*. Elsevier B.V., 69, pp. 96–100. doi: 10.1016/j.jcv.2015.06.082.

Tiffany, A. *et al.* (2017) 'Estimating the number of secondary Ebola cases resulting from an unsafe burial and risk factors for transmission during the West Africa Ebola epidemic', *PLoS Negl Trop Dis.*, 11(6), pp. 1–15.

Toan, N. L. *et al.* (2013) 'Co-infection of human parvovirus B19 with Plasmodium falciparum contributes to malaria disease severity in Gabonese patients.', *BMC infectious diseases*. BMC Infectious Diseases, 13(1), p. 375. doi: 10.1186/1471-2334-13-375.

Towner, J. S. *et al.* (2004) 'Rapid Diagnosis of Ebola Hemorrhagic Fever by Reverse Transcription-PCR in an Outbreak Setting and Assessment of Patient Viral Load as a

Predictor of Outcome', *Journal of virology*, 78(8), pp. 4330–4341. doi: 10.1128/JVI.78.8.4330.

UNAIDS (2016) *Overview of Guinea*. Available at: https://www.unaids.org/en/regionscountries/countries/guinea (Accessed: 5 January 2019).

Uyeki, T. M. *et al.* (2016) 'Clinical management of Ebola virus disease in the United States and Europe', *N Engl J Med.*, 374. doi: 10.1056/NEJMoa1504874.

Varkey, J. B. *et al.* (2015) 'Persistence of Ebola Virus in Ocular Fluid during Convalescence', *New England Journal of Medicine*, 372(25), pp. 2423–2427. doi: 10.1056/NEJMoa1500306.

Vega, M.-A. de La *et al.* (2015) 'Ebola viral load at diagnosis associates with patient outcome and outbreak evolution', *Journal of Clinical Investigation*, 125(12), pp. 4421–4428. doi: 10.1172/jci83162.

Villadangos, J. A. and Shortman, K. (2010) 'Found in translation: the human equivalent of mouse CD8+ dendritic cells', *J Exp Med.*, 207. doi: 10.1084/jem.20100985.

Villinger, F. *et al.* (1999) 'Markedly elevated levels of interferon (IFN)-gamma, IFN-alpha, interleukin (IL)-2, IL-10, and tumor necrosis factor-alpha associated with fatal Ebola virus infection.', *The Journal of infectious diseases*, 179 Suppl(Il), pp. S188–S191. doi: 10.1086/514283.

Volchkov, V. E. *et al.* (2000) 'Molecular characterization of guinea pig-adapted variants of Ebola virus', *Virology*, 277(1), pp. 147–155. doi: 10.1006/viro.2000.0572.

Wang, X. *et al.* (2010) 'Coinfection with EBV / CMV and other respiratory agents in children with suspected infectious mononucleosis', pp. 10–14.

Wauquier, N. *et al.* (2010) 'Human fatal zaire ebola virus infection is associated with an aberrant innate immunity and with massive lymphocyte apoptosis', *PLoS Neglected*

Tropical Diseases, 4(10). doi: 10.1371/journal.pntd.0000837.

Waxman, M. *et al.* (2017) 'Characteristics and survival of patients with Ebola virus infection, malaria, or both in Sierra Leone: a retrospective cohort study', *The Lancet Infectious Diseases*. Elsevier Ltd, 3099(17), pp. 1–7. doi: 10.1016/S1473-3099(17)30112-3.

Weyer, J., Grobbelaar, A. and Blumberg, L. (2015) 'Ebola Virus Disease: History, Epidemiology and Outbreaks', *Current Infectious Disease Reports*, 17(5), pp. 1–8. doi: 10.1007/s11908-015-0480-y.

Whitmer, S. L. M. *et al.* (2018) 'Active Ebola Virus Replication and Heterogeneous Evolutionary Rates in EVD Survivors', *Cell Reports*, 22(5), pp. 1159–1168. doi: 10.1016/j.celrep.2018.01.008.

WHO (2014) 'Ebola Virus Disease in West Africa - The First 9 Months of the Epidemic and Forward Projections', *The New England journal of Medicine*, 371(16), pp. 1481–1495. doi: 10.1056/NEJMoa1005372.

WHO (2016) *Map of Ebola cases in West Africa from January 2014 to March 2016.* Available at: https://www.who.int/csr/disease/ebola/maps/en/ (Accessed: 4 March 2019).

WHO (2018) *World Malaria Report 2018: Guinea*. Available at: https://www.who.int/malaria/publications/country-profiles/profile_gin_en.pdf?ua=1.

Wick, R. R. *et al.* (2017) 'Completing bacterial genome assemblies with multiplex MinION sequencing', *Microbial Genomics*. Microbiology Society, 3(10), p. e000132. doi: 10.1099/mgen.0.000132.

Williams, K. J. *et al.* (2015) 'VSVDeltaG/EBOV GP-induced innate protection enhances natural killer cell activity to increase survival in a lethal mouse adapted Ebola virus infection', *Viral Immunol.*, 28. doi: 10.1089/vim.2014.0069.

Wolf, T. *et al.* (2015) 'Severe Ebola virus disease with vascular leakage and multiorgan failure: Treatment of a patient in intensive care', *The Lancet*, 385(9976), pp. 1428–1435. doi: 10.1016/S0140-6736(14)62384-9.

Wong, G. *et al.* (2016) 'Pathogenicity Comparison between the Kikwit and Makona Ebola Virus Variants in Rhesus Macaques', *Journal of Infectious Diseases*, 214(Suppl 3), pp. S281–S289. doi: 10.1093/infdis/jiw267.

Wong, G., Kobinger, G. P. and Qiu, X. (2014) 'Characterization of host immune responses in Ebola virus infections.', *Expert review of clinical immunology*, 10(6), pp. 781–790. doi: 10.1586/1744666X.2014.908705.

Woolhouse, M. E. J. (1998) 'Patterns in parasite epidemiology: The peak shift', *Parasitology Today*, 14(10), pp. 428–434. doi: 10.1016/S0169-4758(98)01318-0.

Zahn, R. C. *et al.* (2007) 'A-to-G Hypermutation in the Genome of Lymphocytic Choriomeningitis Virus', *Journal of Virology*, 81(2), pp. 457–464. doi: 10.1128/jvi.00067-06.

Zampieri, C. A., Sullivan, N. J. and Nabel, G. J. (2007) 'Immunopathology of highly virulent pathogens: insights from Ebola virus.', *Nature immunology*, 8(11), pp. 1159–1164. doi: 10.1038/ni1519.

Zhang, A. P. P. *et al.* (2012) 'The ebola virus interferon antagonist VP24 directly binds STAT1 and has a novel, pyramidal fold', *PLoS Pathogens*, 8(2). doi: 10.1371/journal.ppat.1002550.

Zumbrun, E. E. *et al.* (2012) 'A Characterization of Aerosolized Sudan Virus Infection in African Green Monkeys, Cynomolgus Macaques, and Rhesus Macaques', *Viruses*, pp. 2115–2136. doi: 10.3390/v4102115.

Chapter 7: Appendix

7.1 Differentially expressed genes following EdgeR analysis

Table 7.1: Top 20 differentially expressed genes of hospitalised fatalities compared to convalescent controls.

Ensembl gene ID	logFC	logCPM	P value	FDR	Gene name	Description
ENSG00000171560	14.3227086	9.43926688	2.05E-09	5.56E-08	FGA	fibrinogen alpha chain [Source:HGNC Symbol;Acc:HGNC:3661]
ENSG00000163631	12.5770161	8.25071724	1.71E-10	6.07E-09	ALB	albumin [Source:HGNC Symbol;Acc:HGNC:399]
ENSG00000171564	12.5102601	7.62897412	2.54E-11	1.07E-09	FGB	fibrinogen beta chain [Source:HGNC Symbol;Acc:HGNC:3662]
ENSG00000171557	11.4745706	6.59726642	4.43E-10	1.39E-08	FGG	fibrinogen gamma chain [Source:HGNC Symbol;Acc:HGNC:3694]
ENSG00000196616	9.79327839	4.9352305	2.92E-10	9.69E-09	ADH1B	alcohol dehydrogenase 1B (class I), beta polypeptide [Source:HGNC Symbol;Acc:HGNC:250]
ENSG00000084674	9.76055663	4.88601175	5.87E-12	2.80E-10	APOB	apolipoprotein B [Source:HGNC Symbol;Acc:HGNC:603]
ENSG00000060566	9.06228077	5.99590418	1.48E-09	4.14E-08	CREB3L3	cAMP responsive element binding protein 3 like 3 [Source:HGNC Symbol;Acc:HGNC:18855]
ENSG00000112936	8.39231342	3.5083034	4.47E-10	1.40E-08	C7	complement C7 [Source:HGNC Symbol;Acc:HGNC:1346]
ENSG00000170323	8.08361154	3.24208624	2.71E-13	1.77E-11	FABP4	fatty acid binding protein 4 [Source:HGNC Symbol;Acc:HGNC:3559]
ENSG00000107562	7.73398284	4.302603	6.73E-12	3.19E-10	CXCL12	C-X-C motif chemokine ligand 12 [Source:HGNC Symbol;Acc:HGNC:10672]
ENSG00000185551	7.68479092	3.92808578	1.49E-12	8.07E-11	NR2F2	nuclear receptor subfamily 2 group F member 2 [Source:HGNC Symbol;Acc:HGNC:7976]
ENSG00000100234	7.56508327	4.05098287	1.53E-15	1.33E-13	TIMP3	TIMP metallopeptidase inhibitor 3 [Source:HGNC Symbol;Acc:HGNC:11822]
ENSG00000162692	7.55358865	3.3936028	1.21E-	2.68E-	VCAM1	vascular cell adhesion molecule 1 [Source:HGNC Symbol;Acc:HGNC:12663]

			08	07		
ENSG00000128052	7.4139009	2.58488318	2.06E-13	1.36E-11	KDR	kinase insert domain receptor [Source:HGNC Symbol;Acc:HGNC:6307]
ENSG00000162618	7.27969672	2.50513531	1.46E-12	7.96E-11	ADGRL4	adhesion G protein-coupled receptor L4 [Source:HGNC Symbol;Acc:HGNC:20822]
ENSG00000139209	7.26202484	5.59208214	1.18E-06	1.64E-05	SLC38A4	solute carrier family 38 member 4 [Source:HGNC Symbol;Acc:HGNC:14679]
ENSG00000187608	7.24116056	5.53064327	4.22E-06	5.19E-05	ISG15	ISG15 ubiquitin-like modifier [Source:HGNC Symbol;Acc:HGNC:4053]
ENSG00000142748	7.24005124	2.32058846	6.80E-11	2.62E-09	FCN3	ficolin 3 [Source:HGNC Symbol;Acc:HGNC:3625]
ENSG00000163958	7.19189378	3.63965586	1.20E-14	9.57E-13	ZDHHC19	zinc finger DHHC-type containing 19 [Source:HGNC Symbol;Acc:HGNC:20713]
ENSG00000168309	7.176219	3.25834384	4.28E-12	2.08E-10	FAM107A	family with sequence similarity 107 member A [Source:HGNC Symbol;Acc:HGNC:30827]

Table 7.2: Top 20 DE genes in hospitalised survivors compared to fatalities.

Ensembl gene ID	logFC	logCPM	P value	FDR	Gene name	Description
ENSG00000276197	-5.3367384	5.02321837	1.32E-13	4.56E-11	pRNA	NoRC assoicated RNA, pRNA [Source:RFAM;Acc:RF01518]
ENSG00000188257	-4.6144515	3.26684373	1.98E-10	2.72E-08	PLA2G2A	phospholipase A2 group IIA [Source:HGNC Symbol;Acc:HGNC:9031]
ENSG00000104760	-4.2386355	6.84742069	4.92E-06	0.00014445	FGL1	fibrinogen like 1 [Source:HGNC Symbol;Acc:HGNC:3695]
ENSG00000171564	-4.1642728	7.16264872	1.24E-08	9.26E-07	FGB	fibrinogen beta chain [Source:HGNC Symbol;Acc:HGNC:3662]
ENSG00000000971	-4.0553824	7.34331921	1.21E-13	4.27E-11	CFH	complement factor H [Source:HGNC Symbol;Acc:HGNC:4883]
ENSG00000171560	-3.8737645	8.98353936	1.16E-05	0.00029276	FGA	fibrinogen alpha chain [Source:HGNC Symbol;Acc:HGNC:3661]
ENSG00000060566	-3.8391213	5.54806362	6.68E-10	8.22E-08	CREB3L3	cAMP responsive element binding protein 3 like 3 [Source:HGNC Symbol;Acc:HGNC:18855]
ENSG00000196616	-3.6651907	4.51766564	4.46E-08	2.74E-06	ADH1B	alcohol dehydrogenase 1B (class I), beta polypeptide [Source:HGNC Symbol;Acc:HGNC:250]
ENSG00000182326	-3.6569443	4.24741588	2.14E-09	2.04E-07	C1S	complement C1s [Source:HGNC Symbol;Acc:HGNC:1247]
ENSG00000091181	-3.5125941	4.76118365	3.46E-11	6.13E-09	IL5RA	interleukin 5 receptor subunit alpha [Source:HGNC Symbol;Acc:HGNC:6017]
ENSG00000139209	-3.4273524	5.17321021	3.22E-07	1.44E-05	SLC38A4	solute carrier family 38 member 4 [Source:HGNC Symbol;Acc:HGNC:14679]
ENSG00000163631	-3.3874133	7.8196547	3.91E-06	0.0001198	ALB	albumin [Source:HGNC Symbol;Acc:HGNC:399]
ENSG00000106853	-3.3399954	4.57266783	7.41E-10	8.75E-08	PTGR1	prostaglandin reductase 1 [Source:HGNC Symbol;Acc:HGNC:18429]
ENSG00000166741	-3.3337968	2.53225277	1.81E-09	1.78E-07	NNMT	nicotinamide N-methyltransferase [Source:HGNC Symbol;Acc:HGNC:7861]
ENSG00000171557	-	6.17417424	5.75E-06	0.00016442	FGG	fibrinogen gamma chain [Source:HGNC Symbol;Acc:HGNC:3694]

	3.2981538					
ENSG00000154736	-3.2737389	1.89119184	1.36E-09	1.41E-07	ADAMTS5	ADAM metallopeptidase with thrombospondin type 1 motif 5 [Source:HGNC Symbol;Acc:HGNC:221]
ENSG00000131462	-3.2054854	2.84530408	2.63E-08	1.76E-06	TUBG1	tubulin gamma 1 [Source:HGNC Symbol;Acc:HGNC:12417]
ENSG00000276700	-3.0669568	7.0927172	6.50E-08	3.76E-06	RNA5-8SN1	RNA, 5.8S ribosomal N1 [Source:HGNC Symbol;Acc:HGNC:53517]
ENSG00000162692	-3.0575901	4.45658564	1.60E-15	7.25E-13	VCAM1	vascular cell adhesion molecule 1 [Source:HGNC Symbol;Acc:HGNC:12663]
ENSG00000187608	-3.0572754	3.02647961	4.77E-07	2.03E-05	ISG15	ISG15 ubiquitin-like modifier [Source:HGNC Symbol;Acc:HGNC:4053]

7.2 KEGG pathway analysis

Table 7.3: KEGG analysis enriched pathways of hospitalised fatalities compared to convalescent controls.

ID	Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue
hsa05144	Malaria	18/660	49/7383	8.63E-08	2.60E-05	2.23E-05
hsa04610	Complement and coagulation cascades	23/660	79/7383	2.05E-07	3.08E-05	2.64E-05
hsa05322	Systemic lupus erythematosus	30/660	133/7383	1.42E-06	0.00014255	0.00012214
hsa04668	TNF signaling pathway	26/660	108/7383	1.97E-06	0.00014791	0.00012673
hsa04060	Cytokine-cytokine receptor interaction	45/660	270/7383	2.60E-05	0.00156802	0.00134347
hsa04514	Cell adhesion molecules (CAMs)	28/660	145/7383	6.78E-05	0.00340164	0.0029145
hsa04640	Hematopoietic cell lineage	20/660	97/7383	0.00029594	0.01013005	0.00867936
hsa05215	Prostate cancer	20/660	97/7383	0.00029594	0.01013005	0.00867936
hsa05150	Staphylococcus aureus infection	14/660	56/7383	0.00030289	0.01013005	0.00867936
hsa05202	Transcriptional misregulation in cancer	31/660	186/7383	0.0004668	0.01289018	0.01104422
hsa03320	PPAR signaling pathway	16/660	72/7383	0.00048939	0.01289018	0.01104422
hsa05418	Fluid shear stress and atherosclerosis	25/660	139/7383	0.00051389	0.01289018	0.01104422
hsa05133	Pertussis	16/660	76/7383	0.00091879	0.02127354	0.01822702
hsa05167	Kaposi's sarcoma-associated herpesvirus infection	30/660	186/7383	0.00099822	0.02146182	0.01838834
hsa05216	Thyroid cancer	10/660	37/7383	0.00114164	0.02290891	0.0196282
hsa04064	NF-kappa B signaling pathway	18/660	95/7383	0.00164551	0.0309561	0.02652297
hsa04015	Rap1 signaling pathway	31/660	206/7383	0.00259247	0.04590195	0.03932848
hsa05143	African trypanosomiasis	9/660	35/7383	0.00291415	0.04873102	0.04175241

Table 7.4: KEGG analysis enriched pathways in hospitalised fatalities compared to survivors aged 5-45.

ID	Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue
hsa04610	Complement and coagulation cascades	16/75	79/7383	1.08E-11	2.49E-09	2.08E-09
hsa04060	Cytokine-cytokine receptor interaction	11/75	270/7383	7.80E-05	0.00663011	0.00595273
hsa05143	African trypanosomiasis	4/75	35/7383	0.00040521	0.01750511	0.01571666
hsa05142	Chagas disease (American trypanosomiasis)	6/75	102/7383	0.00055896	0.01750511	0.01571666
hsa04668	TNF signalling pathway	7/75	108/7383	0.0006178	0.01221869	0.00989923
hsa04620	Toll-like receptor signalling pathway	6/75	104/7383	0.00061988	0.01750511	0.01571666
hsa04622	RIG-I-like receptor signalling pathway	5/75	70/7383	0.00068442	0.01750511	0.01571666
hsa04659	Th17 cell differentiation	6/75	107/7383	0.0007208	0.01750511	0.01571666
hsa05133	Pertussis	5/75	76/7383	0.00099657	0.02117709	0.01901349
hsa05144	Malaria	4/75	49/7383	0.00147289	0.0278213	0.02497888
hsa04064	NF-kappa B signaling pathway	7/75	95/7383	0.00174639	0.02124589	0.01721282
Hsa04062	Chemokine signalling pathway	7/75	133/7383	0.00080803	0.02383708	0.02147675
hsa04650	Natural killer cell mediated cytotoxicity	6/75	133/7383	0.00222197	0.03777347	0.03391426

Table 7.5: KEGG analysis enriched pathways in hospitalised fatalities compared to survivors aged > 45 years.

ID	Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue
hsa04610	Complement and coagulation cascades	13/149	79/7383	4.58E-09	1.08E-06	9.73E-07
hsa05205	Proteoglycans in cancer	16/149	201/7383	2.53E-06	0.0002988	0.00026921
hsa05150	Staphylococcus aureus infection	7/149	56/7383	0.00011997	0.00943738	0.0085029
hsa04010	MAPK signalling pathway	16/149	295/7383	0.00028057	0.0165535	0.01491439
hsa05202	Transcriptional misregulation in cancer	12/149	186/7383	0.00035812	0.01690324	0.0152295
hsa04979	Cholesterol metabolism	6/149	50/7383	0.00046672	0.01835782	0.01654006
hsa04014	Ras signalling pathway	13/149	232/7383	0.00078736	0.02383708	0.02147676
hsa04060	Cytokine-cytokine receptor interaction	7/149	76/7383	0.00080804	0.02383708	0.02147676

Table 7.6: KEGG analysis enriched pathways in hospitalised fatalities compared to survivors aged < 5 years.

ID	Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue
hsa05322	Systemic lupus erythematosus	3/5	133/7383	5.57E-05	0.00027835	5.86E-05
hsa05144	Malaria	3/5	180/7383	0.00013749	0.00034372	7.24E-05
hsa04217	Necroptosis	2/5	164/7383	0.00469303	0.00636714	0.00134045
hsa03013	RNA transport	2/5	171/7383	0.00509371	0.00636714	0.00134045

7.3 PCR Profiler Array: Human Inflammatory Cytokines and Receptors (Qiagen)

Table 7.7: Gene table RT2 Profiler PCR Array: Human Inflammatory Cytokines and Receptors. *96-well position shown for 84 target genes shown along with housekeeping genes and controls.*

Position	UniGene	GenBank	Symbol	Description
A01	Hs.591680	NM_004757	AIMP1	Aminoacyl tRNA synthetase complex-interacting multifunctional protein 1
A02	Hs.73853	NM_001200	BMP2	Bone morphogenetic protein 2
A03	Hs.494997	NM_001735	C5	Complement component 5
A04	Hs.72918	NM_002981	CCL1	Chemokine (C-C motif) ligand 1
A05	Hs.54460	NM_002986	CCL11	Chemokine (C-C motif) ligand 11
A06	Hs.414629	NM_005408	CCL13	Chemokine (C-C motif) ligand 13
A07	Hs.272493	NM_032965	CCL15	Chemokine (C-C motif) ligand 15
A08	Hs.10458	NM_004590	CCL16	Chemokine (C-C motif) ligand 16
A09	Hs.546294	NM_002987	CCL17	Chemokine (C-C motif) ligand 17
A10	Hs.303649	NM_002982	CCL2	Chemokine (C-C motif) ligand 2
A11	Hs.75498	NM_004591	CCL20	Chemokine (C-C motif) ligand 20
A12	Hs.534347	NM_002990	CCL22	Chemokine (C-C motif) ligand 22
B01	Hs.169191	NM_005064	CCL23	Chemokine (C-C motif) ligand 23
B02	Hs.247838	NM_002991	CCL24	Chemokine (C-C motif) ligand 24
B03	Hs.131342	NM_006072	CCL26	Chemokine (C-C motif) ligand 26
B04	Hs.514107	NM_002983	CCL3	Chemokine (C-C motif) ligand 3
B05	Hs.75703	NM_002984	CCL4	Chemokine (C-C motif) ligand 4
B06	Hs.514821	NM_002985	CCL5	Chemokine (C-C motif) ligand 5
B07	Hs.251526	NM_006273	CCL7	Chemokine (C-C motif) ligand 7
B08	Hs.271387	NM_005623	CCL8	Chemokine (C-C motif) ligand 8
B09	Hs.301921	NM_001295	CCR1	Chemokine (C-C motif) receptor 1
B10	Hs.511794	NM_001123396	CCR2	Chemokine (C-C motif) receptor 2
B11	Hs.506190	NM_001837	CCR3	Chemokine (C-C motif) receptor 3
B12	Hs.184926	NM_005508	CCR4	Chemokine (C-C motif) receptor 4
C01	Hs.450802	NM_000579	CCR5	Chemokine (C-C motif) receptor 5
C02	Hs.46468	NM_004367	CCR6	Chemokine (C-C motif) receptor 6
C03	Hs.113222	NM_005201	CCR8	Chemokine (C-C motif) receptor 8
C04	Hs.592244	NM_000074	CD40LG	CD40 ligand
C05	Hs.591402	NM_000757	CSF1	Colony stimulating factor 1 (macrophage)
C06	Hs.1349	NM_000758	CSF2	Colony stimulating factor 2 (granulocyte-macrophage)
C07	Hs.2233	NM_000759	CSF3	Colony stimulating factor 3 (granulocyte)
C08	Hs.531668	NM_002996	CX3CL1	Chemokine (C-X3-C motif) ligand 1
C09	Hs.78913	NM_001337	CX3CR1	Chemokine (C-X3-C motif) receptor 1
C10	Hs.789	NM_001511	CXCL1	Chemokine (C-X-C motif) ligand 1 (melanoma growth stimulating activity, alpha)
C11	Hs.632586	NM_001565	CXCL10	Chemokine (C-X-C motif) ligand 10
C12	Hs.632592	NM_005409	CXCL11	Chemokine (C-X-C motif) ligand 11
D01	Hs.522891	NM_000609	CXCL12	Chemokine (C-X-C motif) ligand 12
D02	Hs.100431	NM_006419	CXCL13	Chemokine (C-X-C motif) ligand 13
D03	Hs.590921	NM_002089	CXCL2	Chemokine (C-X-C motif) ligand 2
D04	Hs.89690	NM_002090	CXCL3	Chemokine (C-X-C motif) ligand 3
D05	Hs.89714	NM_002994	CXCL5	Chemokine (C-X-C motif) ligand 5
D06	Hs.164021	NM_002993	CXCL6	Chemokine (C-X-C motif) ligand 6 (granulocyte chemotactic protein 2)
D07	Hs.77367	NM_002416	CXCL9	Chemokine (C-X-C motif) ligand 9
D08	Hs.194778	NM_000634	CXCR1	Chemokine (C-X-C motif) receptor 1
D09	Hs.846	NM_001557	CXCR2	Chemokine (C-X-C motif) receptor 2

Position	UniGene	GenBank	Symbol	Description
D10	Hs.2007	NM_000639	FASLG	Fas ligand (TNF superfamily, member 6)
D11	Hs.211575	NM_000605	IFNA2	Interferon, alpha 2
D12	Hs.856	NM_000619	IFNG	Interferon, gamma
E01	Hs.504035	NM_001558	IL10RA	Interleukin 10 receptor, alpha
E02	Hs.654593	NM_000628	IL10RB	Interleukin 10 receptor, beta
E03	Hs.845	NM_002188	IL13	Interleukin 13
E04	Hs.654378	NM_000585	IL15	Interleukin 15
E05	Hs.459095	NM_004513	IL16	Interleukin 16
E06	Hs.41724	NM_002190	IL17A	Interleukin 17A
E07	Hs.278911	NM_013278	IL17C	Interleukin 17C
E08	Hs.272295	NM_052872	IL17F	Interleukin 17F
E09	Hs.1722	NM_000575	IL1A	Interleukin 1, alpha
E10	Hs.126256	NM_000576	IL1B	Interleukin 1, beta
E11	Hs.701982	NM_000877	IL1R1	Interleukin 1 receptor, type I
E12	Hs.81134	NM_000577	IL1RN	Interleukin 1 receptor antagonist
F01	Hs.567559	NM_021803	IL21	Interleukin 21
F02	Hs.528111	NM_145659	IL27	Interleukin 27
F03	Hs.694	NM_000588	IL3	Interleukin 3 (colony-stimulating factor, multiple)
F04	Hs.348390	NM_033439	IL33	Interleukin 33
F05	Hs.2247	NM_000879	IL5	Interleukin 5 (colony-stimulating factor, eosinophil)
F06	Hs.68876	NM_000564	IL5RA	Interleukin 5 receptor, alpha
F07	Hs.591873	NM_000880	IL7	Interleukin 7
F08	Hs.624	NM_000584	IL8	Interleukin 8
F09	Hs.960	NM_000590	IL9	Interleukin 9
F10	Hs.406228	NM_002186	IL9R	Interleukin 9 receptor
F11	Hs.36	NM_000595	LTA	Lymphotoxin alpha (TNF superfamily, member 1)
F12	Hs.376208	NM_002341	LTB	Lymphotoxin beta (TNF superfamily, member 3)
G01	Hs.407995	NM_002415	MIF	Macrophage migration inhibitory factor (glycosylation-inhibiting factor)
G02	Hs.489615	NM_005746	NAMPT	Nicotinamide phosphoribosyltransferase
G03	Hs.248156	NM_020530	OSM	Oncostatin M
G04	Hs.313	NM_000582	SPP1	Secreted phosphoprotein 1
G05	Hs.241570	NM_000594	TNF	Tumor necrosis factor
G06	Hs.81791	NM_002546	TNFRSF11B	Tumor necrosis factor receptor superfamily, member 11b
G07	Hs.478275	NM_003810	TNFSF10	Tumor necrosis factor (ligand) superfamily, member 10
G08	Hs.333791	NM_003701	TNFSF11	Tumor necrosis factor (ligand) superfamily, member 11
G09	Hs.54673	NM_003808	TNFSF13	Tumor necrosis factor (ligand) superfamily, member 13
G10	Hs.525157	NM_006573	TNFSF13B	Tumor necrosis factor (ligand) superfamily, member 13b
G11	Hs.181097	NM_003326	TNFSF4	Tumor necrosis factor (ligand) superfamily, member 4
G12	Hs.73793	NM_003376	VEGFA	Vascular endothelial growth factor A
H01	Hs.520640	NM_001101	ACTB	Actin, beta
H02	Hs.534255	NM_004048	B2M	Beta-2-microglobulin
H03	Hs.592355	NM_002046	GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
H04	Hs.412707	NM_000194	HPRT1	Hypoxanthine phosphoribosyltransferase 1
H05	Hs.546285	NM_001002	RPLP0	Ribosomal protein, large, P0
H06	N/A	SA_00105	HGDC	Human Genomic DNA Contamination
H07	N/A	SA_00104	RTC	Reverse Transcription Control
H08	N/A	SA_00104	RTC	Reverse Transcription Control
H09	N/A	SA_00104	RTC	Reverse Transcription Control
H10	N/A	SA_00103	PPC	Positive PCR Control
H11	N/A	SA_00103	PPC	Positive PCR Control
H12	N/A	SA_00103	PPC	Positive PCR Control